



Universidad
Carlos III de Madrid

TESIS DOCTORAL

Modelado Ubicuo del Comportamiento para la Asistencia de Personas Mayores en el Hogar

Autor:

Francisco Javier Ordóñez Morales

Directores:

Paula de Toledo Heras

Araceli Sanchis de Miguel

DEPARTAMENTO DE INFORMÁTICA

Leganés, Septiembre 2013

Tribunal nombrado por el Mgfco. y Excmo. Sr. Rector de la Universidad Carlos III de Madrid, el día de de 2013.

Presidente: D.

Vocal: D.

Secretario: D.

Realizado el acto de defensa y lectura de la Tesis el día de de 2013 en

Calificación:

EL PRESIDENTE

LOS VOCALES

EL SECRETARIO

A Eva

“La utopía está en el horizonte, yo sé muy bien que nunca la alcanzaré, que si yo camino diez pasos, ella se aleja diez pasos y cuanto más la busque... menos la encontraré. Entonces, ¿para qué sirve la utopía?. Pues la utopía sirve para eso, para caminar.”

Eduardo Galeano

Agradecimientos

Un paso más. De un camino que jamás podría recorrer si no fuera por todas las personas que me he encontrado en él, todas esas personas que me han servido de ayuda y reposo, a las cuales estoy profundamente agradecido.

Cuando pienso en a quien dar las gracias, las primeras personas en las que puedo pensar son mis padres. Su apoyo incondicional ha sido clave en muchos momentos de mi vida, y este no es una excepción. Gracias.

A mis directoras de tesis, Paula y Araceli, por su apoyo constante, consejos y tiempo. Ha sido un placer poder trabajar y aprender con ellas.

A mi hermana Maria Jesus, que no ha dudado en venir cuando la he necesitado, aunque para ello tuviese que cruzar fronteras.

A mi hermano Oscar, que tanto me ha ayudado cuando me quedaba atascado con la electrónica y que incluso vino conmigo a instalar la instrumentación cuando me faltaron manos, si he sido capaz de hacer funcionar uno sólo de los sensores, ha sido gracias a él.

A la sección holandesa de este trabajo, a Ben Kröse, Gwenn Englebienne, Bas Terwijn y Tim van Kasteren, cuyos valiosísimos consejos y recomendaciones fueron claves en los inicios de esta tesis.

También quiero agradecer a las organizaciones que de una forma u otra han apoyado la realización de esta tesis doctoral, a la Asociación Párkinson Madrid y al programa AAL, el cual ha financiado gran parte de este trabajo. A Amelia y a Juli, por su tiempo, colaboración y paciencia.

A todos mis compañeros del grupo CAOS, los presentes y pasados. A Agapito, por enseñarme a dar mis primeros pasos en el mundo de la investigación y confiar siempre en mí. A Jorge, por su generosidad con su tiempo y conocimiento. A Paz, por esos consejos necesarios y por todo lo que me ha enseñado sobre la docencia. A Andrés, por su implicación y ayuda en numerosas partes de este trabajo. A Jose Antonio, por su compañía y ayuda desinteresada. A Raúl y JP, por todas esas sonrisas y entretenidos descansos. Al resto de miembros del grupo que han ayudado siempre a crear un extraordinario ambiente de trabajo, Beatriz, Juan,

X

Germán, Isaac, Ariel y Carlos.

Y por último, dar las gracias a una persona muy especial, a Eva. Mi redactora, correctora, traductora, editora, ayudante, secretaria, psicóloga, consejera y por encima de todo, mi compañera.

Gracias a todos por haber servido de apoyo en mi camino. Seguiremos dando pasos, persiguiendo el horizonte, aunque no para alcanzar la utopía, sino para no dejar nunca de caminar.

Resumen

El envejecimiento de la población está teniendo un impacto muy relevante en los sistemas de salud en todo el mundo. Debido a esta situación, hoy día existe un creciente interés en el desarrollo de sistemas automáticos de supervisión y alarma que faciliten vivir de forma independiente a las personas mayores durante el mayor tiempo posible. Los sistemas de inteligencia ambiental proporcionan una manera sencilla de mejorar la sensación de seguridad de las personas que viven de forma independiente, y permiten a las personas mayores ser más autosuficientes, fomentando su autonomía.

Unos indicadores muy comunes para evaluar el bienestar físico y cognitivo de las personas mayores son los cambios o desviaciones en sus Actividades de Vida Diaria (AVDs) y en sus patrones de comportamiento. El conjunto de actividades que se conocen como AVDs son aquellas que se llevan a cabo con una periodicidad diaria y se consideran definitorias del comportamiento general de un individuo. Dentro de esta definición se incluyen actividades tales como asearse, alimentarse o bañarse. Recientes desarrollos relativos a las tecnologías de monitorización hacen posible instrumentar cualquier entorno de un modo sencillo y barato, ofreciendo un mecanismo para poder controlar en todo momento la situación de una vivienda real.

En esta tesis doctoral se ha hecho uso de redes de sensores inalámbricos para registrar el comportamiento de las personas en su propia vivienda. Empleando sensores digitales se han controlando tanto el movimiento de los usuarios, como el uso que hacían de electrodomésticos, muebles o puertas. Modelar el comportamiento humano no es un problema trivial, debido a que los datos generados por los sensores son generalmente ruidosos y tienen un comportamiento no determinista, y además cada usuario realiza cada actividad específica de un modo personal y diferente. Modelos probabilísticos clásicos basados en el algoritmo de estimación de máxima verosimilitud han demostrado ofrecer un buen comportamiento en este dominio, sin embargo, dichos modelos son propensos a sobreadaptarse a los datos y generalmente requieren de la aceptación de suposiciones muy estrictas acerca de la distribución de los atributos.

En esta tesis doctoral, se aborda el problema de modelar el comportamiento

humano a través de dos enfoques de aprendizaje diferentes. Se propone un sistema para el reconocimiento de las AVDs mediante el uso de aprendizaje supervisado, y un método de detección de anomalías que identifica estadísticamente patrones de comportamiento atípicos de un modo no supervisado.

Para abordar el problema del reconocimiento de actividades, se presentan dos propuestas basadas en los modelos ocultos de Markov (HMM), a saber: esquemas HMM híbridos y un método de inferencia bayesiano para entrenar un HMM estándar. Un HMM es un modelo probabilístico temporal que puede ser empleado para reconocer actividades humanas de un modo eficaz, pero que requiere asumir condiciones de independencia muy estrictas entre los atributos, que muy probablemente no representan la distribución real de los datos. En este trabajo se muestra cómo la combinación de las capacidades discriminativas de clasificadores bien conocidos, como son las redes de neuronas y las máquinas de vectores de soporte, junto a la capacidad de los HMM de capturar las relaciones temporales de los datos, ofrece una mejor tasa de reconocimiento y un marco más flexible para modelar el espacio de características de los sensores.

También se demuestra que las técnicas de simulación estocástica, como el método Montecarlo basado en cadenas de Markov (MCMC), pueden ser empleadas para estimar eficazmente los parámetros de un HMM haciendo uso de estadística bayesiana. Se evitan además los problemas que típicamente acompañan a los métodos clásicos de inferencia basados en máxima verosimilitud, a saber: son numéricamente sensibles, dependen fuertemente de las estimaciones iniciales de los parámetros y su tipo de búsqueda suele conducir a soluciones subóptimas del problema.

En lo relativo al método propuesto para la detección no supervisada de anomalías en los patrones del comportamiento, se presenta un algoritmo basado en estadística bayesiana que se puede emplear para identificar comportamientos atípicos en una persona de forma eficaz. Los patrones de comportamiento de las personas se modelan estadísticamente mediante tres probabilidades que presentamos en este trabajo. Dichas probabilidades son definitorias de: (1) cuándo se activan los sensores, (2) en qué orden y (3) durante cuánto tiempo. El método se basa en una estimación aproximada de las métricas que definen los patrones diarios de las personas, y en unas probabilidades a priori que reflejan un conocimiento previo acerca de cómo deben ser dichas métricas. Estas estimaciones se emplean para detectar signos anómalos en el comportamiento que pudieran indicar un cambio en la salud del usuario.

Finalmente, se detalla una propuesta para aplicar los métodos de modelado del comportamiento presentados en múltiples entornos monitorizados, sin necesidad de obtener datos de entrenamiento de cada entorno particular. Tanto los métodos supervisados como los no supervisados, cuando se optimizan para una vivienda, no pueden ser automáticamente empleados para otra vivienda distinta, debido a

las diferencias tanto entre el diseño de casas como en el comportamiento de sus habitantes. En esta tesis se presenta un método de transferencia de información para los problemas de reconocimiento de actividades y detección de anomalías que permite emplear el conocimiento adquirido en diferentes viviendas bajo un enfoque bayesiano, y sirve como base para la instalación de los sistemas en un entorno monitorizado completamente nuevo, ofreciendo una solución para aplicar los métodos a mayor escala.

En este trabajo de tesis doctoral, se evalúan las diferentes propuestas usando tres conjuntos de datos generados por el sistema de monitorización presentado, y otros tres conjuntos de datos públicos, ampliamente usados en la literatura. Los resultados experimentales han demostrado cómo los modelos HMM híbridos pueden reconocer eficazmente las actividades del usuario, mejorando significativamente tanto a su versión puramente discriminativa como a los modelos generativos. También se ha demostrado cómo el rendimiento de un HMM puede mejorar significativamente si se emplea inferencia bayesiana bajo un algoritmo MCMC. Este método ofrece una búsqueda mucho más completa en el espacio de soluciones, y permite integrar la estimación de la cadena oculta dentro del proceso de inferencia. Finalmente, los resultados de los experimentos en transferencia de información confirman que existe la posibilidad de diseñar métodos que permitan abstraer las capacidades de detección de un entorno automáticamente, siendo posible tener algoritmos de modelado que puedan funcionar en meta espacios de características, y de este modo permitan trasladar el aprendizaje desde un entorno instrumentado a otro.

Para concluir, esta tesis doctoral confirma que es posible modelar el comportamiento humano de una forma eficaz mediante un conjunto de sensores inalámbricos sencillos y baratos, empleando tanto enfoques supervisados como no supervisados.

Abstract

Population aging is currently having a significant impact on health care systems. As demographics age and the burden of healthcare on society increases, the need for automated, ambient monitoring and alarming systems becomes more predominant. Automated and ambient systems provide an easy way for increasing the safety and the sense of security of people living on their own, and allow the elderly to be self-reliant longer, fostering their autonomy. Common indicators for assessing the cognitive and physical wellbeing of an elderly person are the changes in the Activities of Daily Living (ADLs) and in the behavior patterns. Activities referred to as ADLs are everyday tasks essential to self-care and independent living, such as cooking, toileting and showering. Recent developments in sensing technology make it possible to inexpensively equip existing homes with sensors, therefore allowing a continuous monitoring system.

In this work, we use wireless sensor networks to capture the behavior of the inhabitants in their homes. Digital state-change sensors are used to measure the motion of the inhabitants and the use of appliances, home devices, pieces of furniture and doors. Modeling human behavior from such sensor data is a challenging problem because the data is ambiguous, unsegmented, noisy and also because there are changes in how each person performs a specific action. Probabilistic models using classical, maximum-likelihood estimation methods are known to work well in this domain, but they are prone to overfitting and typically require strong model assumptions, not offering a flexible solution.

In this thesis, we address the human behavior modeling problem through the use of two different learning approaches. We propose an ADLs recognition system based on supervised learning and an anomaly detection approach to statistically identify unusual behavior patterns in an unsupervised way.

In order to address the activity recognition problem, we present two approaches based on hidden Markov models (HMM), namely: hybrid HMM schemes and a standard HMM trained by full bayesian inference. HMM is a temporal probabilistic model that can be effectively used for recognizing human activities, but requires to apply strong model assumptions which most likely does not represent the true distribution of the data, as the complete independence of every sensor.

We showed that the combination of the discriminative capabilities of a machine learning scheme and the superior dynamic time warping abilities of the HMM can offer better recognition performance and provide a more flexible framework to model the sensors feature space. We also show how Markov Chain Monte Carlo (MCMC) techniques can be used to properly estimate the parameters of an HMM in a bayesian configuration, overcoming usual drawbacks posed by classical expectation-maximization algorithms: the use of a recurrent forward-backward algorithm is numerically sensitive, it can get stuck as it approaches a local optima and it can easily overfit the training data.

Regarding our unsupervised anomaly detection proposal, we present a method based on bayesian statistics that can be effectively applied for identifying abnormal human behavior. Behavioral patterns of the residents are statistically estimated based on three probabilistic features that we introduce, namely: sensor activation likelihood, sensor sequence likelihood and sensor duration likelihood. The method relies on an approximate estimation of the living patterns of the user, and on prior knowledge that reflects our belief of how such patterns should be, to detect abnormal behavior signs which could reflect changes in health status of the user.

Finally, we introduce an approach for applying our behavior modeling methods in multiple homes without the need of having training data from each home. Both supervised or unsupervised methods, when trained for one home, can not automatically be used in another home, due to differences in the layout of the places and the behavior of the inhabitants. We present a transfer learning method that allows us to use data from other homes to recognize ADLs and estimate the living patterns of the user in a new home. This makes it possible to apply the models discussed on a large scale, therefore providing a broadly applicable solution to efficient care giving of elderly.

In this thesis, we evaluate our approaches using three datasets recorded by our monitoring systems in real world settings and three datasets publicly available. Experimental results of the hybrid HMM models demonstrate that hybrid schemes can outperform other classical sequential pattern recognition approaches, showing how the combination of discriminative and generative models is more accurate than either of the models on their own. Also, fully bayesian inference using MCMC algorithms has been proved to be a more flexible and accurate technique to estimate the HMM parameters for our domain. In the anomaly detection domain, results show that abnormal behavior signs can be statistically identified based on several probabilistic features and Bayesian statistics provides a very consistent way to reason under the uncertainty of human behavior. Finally, experiments on transfer learning show how there exists the possibility to devise methods that allow to abstract the sensing capabilities of an environment quite automatically, being possible to have behavior modeling systems operating in a meta space that allows

the methods to be translated from one sensing configuration to another.

Moreover, the work presented here further demonstrates that accurate human behavior modeling can be achieved by a set of simple and cheap state-change sensors installed in a wireless network, using both supervised and unsupervised approaches.

Índice general

| | |
|---|-----------|
| 1. Introducción | 1 |
| 1.1. Objetivos | 3 |
| 1.2. Organización de la Memoria | 5 |
| 2. Estado del Arte | 7 |
| 2.1. Tecnologías para la Monitorización Ubicua | 8 |
| 2.1.1. Dispositivos Portátiles | 8 |
| 2.1.2. Videocámaras | 10 |
| 2.1.3. Identificación por Radiofrecuencia | 10 |
| 2.1.4. Redes de Sensores Inalámbricos | 11 |
| 2.2. Reconocimiento de Actividad | 12 |
| 2.3. Detección de Patrones de Actividad | 15 |
| 2.4. Transferencia de Información | 18 |
| 2.5. Conjuntos de Datos y Proyectos Académicos | 20 |
| 2.5.1. Proyectos de Casas Inteligentes | 20 |
| 2.5.2. Conjuntos de Datos Públicos | 22 |
| 2.6. Soluciones Comerciales | 23 |
| 3. Captura y Preprocesado de Datos | 25 |
| 3.1. Red Inalámbrica de Sensores | 25 |
| 3.1.1. Tipos de Sensores Empleados | 26 |
| 3.1.2. Arquitectura del Sistema | 27 |
| 3.1.3. Descripción de las Viviendas | 30 |
| 3.2. Anotación de Actividades | 31 |
| 3.3. Notación y Discretización | 35 |
| 4. Reconocimiento de Actividades | 37 |
| 4.1. Modelos Generativos vs Modelos Discriminativos | 40 |
| 4.2. Modelos Ocultos de Markov | 43 |
| 4.2.1. Inferencia | 45 |
| 4.2.2. Aprendizaje de los Parámetros | 45 |

| | | |
|-----------|---|------------|
| 4.3. | Modelos Híbridos | 46 |
| 4.3.1. | Arquitectura Híbrida para el Reconocimiento de AVDs | 50 |
| 4.3.2. | Evaluación | 57 |
| 4.4. | Inferencia Bayesiana en HMM | 73 |
| 4.4.1. | Simulación Estocástica vía Métodos MCMC | 77 |
| 4.4.2. | Evaluación | 82 |
| 4.5. | Conclusiones | 84 |
| 5. | Detección de Anomalías | 87 |
| 5.1. | Metodología | 93 |
| 5.1.1. | Probabilidad de Activación de los Sensores : PAS | 95 |
| 5.1.2. | Probabilidad de Secuencia de los Sensores : PSS | 97 |
| 5.1.3. | Probabilidad de Duración de los Sensores : PDS | 98 |
| 5.1.4. | Clasificación de las Anomalías | 100 |
| 5.2. | Evaluación | 101 |
| 5.2.1. | Conjuntos de Datos | 101 |
| 5.2.2. | Configuración de la Experimentación | 102 |
| 5.2.3. | Optimización de los Parámetros | 103 |
| 5.2.4. | Resultados | 104 |
| 5.2.5. | Ejemplo de Funcionamiento | 109 |
| 5.2.6. | Comparativa | 110 |
| 5.3. | Conclusiones | 112 |
| 6. | Transferencia de Información | 115 |
| 6.1. | Metodología | 117 |
| 6.1.1. | Mapeado mediante Meta Características | 118 |
| 6.1.2. | Estimación de los Parámetros | 119 |
| 6.2. | Evaluación | 121 |
| 6.2.1. | Reconocimiento de Actividades | 122 |
| 6.2.2. | Detección de Anomalías | 129 |
| 6.3. | Conclusiones | 134 |
| 7. | Conclusions | 137 |
| 7.1. | Conclusions and Contributions | 137 |
| 7.2. | Future Research | 141 |
| 7.3. | Publications | 142 |
| A. | Introducción a los Modelos Ocultos de Markov | 171 |
| A.1. | Definición de <i>Modelos Ocultos de Markov</i> | 171 |
| A.2. | Formalización de los HMM | 172 |
| A.3. | Los problemas fundamentales de un HMM | 173 |

| | |
|---|------------|
| A.4. Verosimilitud de una secuencia de observaciones | 173 |
| A.4.1. Cálculo de la verosimilitud: El algoritmo <i>forward</i> | 174 |
| A.5. La secuencia de estados óptima: El algoritmo de <i>Viterbi</i> | 175 |
| A.6. El entrenamiento de los modelos: El algoritmo <i>Baum-Welch</i> | 177 |
| B. Meta Características por Vivienda | 181 |
| C. Evaluación de la Experiencia de Usuario | 187 |
| C.1. Evaluación Cuantitativa | 187 |
| C.2. Evaluación Cualitativa | 190 |
| C.3. Discusión | 191 |

Índice de figuras

| | |
|---|----|
| 3.1. Nodo inalámbrico RFM DM 1810 (a) y nodo base RFM DM 1810 (b). | 26 |
| 3.2. Esquema de comunicación publicación/suscripción. | 28 |
| 3.3. Diseño de la arquitectura del sistema. | 30 |
| 3.4. Diseño de planta del domicilio empleado para generar el conjunto de datos ‘OrdonezA’. | 31 |
| 3.5. Diseño de planta del domicilio empleado para generar el conjunto de datos ‘OrdonezB’. | 32 |
| 3.6. Diseño de planta del domicilio empleado para generar el conjunto de datos ‘OrdonezC’. | 32 |
| 3.7. Segmentación de los datos. | 35 |
| 4.1. Ejemplos de asunciones incorrectas sobre el modelo propuesto por Kasteren <i>et. al.</i> [129]. | 41 |
| 4.2. Representación gráfica de un modelo oculto de Markov. Los nodos y representan los estados ocultos y los nodos \vec{x} las observaciones. | 43 |
| 4.3. Ejemplos de hiperplanos para clasificación binaria. | 52 |
| 4.4. Estructura del modelo híbrido HMM/RNA empleado. | 55 |
| 4.5. Esquema genérico para el procesamiento de nuestros datos en un modelo HMM híbrido. | 56 |
| 4.6. Tipos de representación de los datos. | 59 |
| 4.7. Valores de <i>F-Measure</i> para los modelos discriminativos en cada conjunto de datos, usando diferentes longitudes de ventana deslizante. | 66 |
| 4.8. Valor promediado de <i>F-Measure</i> para todos los conjuntos de datos empleados usando los modelos discriminativos y diferentes longitudes de ventana deslizante. | 67 |
| 4.9. Comparativa promediada para todos los modelos empleados en el análisis de métodos híbridos. | 70 |

| | |
|--|-----|
| 4.10. Propuesta bayesiana para estimar la variable oculta y , a partir de una variable observable x . (a) Distribución a posteriori del parámetro θ dado x . Probabilidad para $y = 1$ dado las observaciones x y el parámetro θ | 75 |
| 5.1. Probabilidad de activación de un sensor instalado en el lavabo empleando estadística frecuentista (línea continua) y nuestra probabilidad (PAS) (línea punteada). | 96 |
| 5.2. Secuencia S de activaciones de los sensores durante un día de datos para el conjunto ‘OrdonezA’. | 98 |
| 5.3. Curva ROC del modelo para el conjunto de datos ‘OrdonezA’. . . | 107 |
| 5.4. Curva ROC del modelo para el conjunto de datos ‘OrdonezB’. . . | 108 |
| 5.5. Curva ROC del modelo para el conjunto de datos ‘OrdonezC’. . . | 108 |
| 5.6. Ejemplo de rendimiento para la métrica PAS en el conjunto de datos ‘OrdonezB’ usando anomalías reales. Los espacios sombreados representan intervalos donde comportamientos anómalos fueron detectados. | 111 |
| 6.1. Esquema de la propuesta para realizar la transferencia de información. | 120 |
| 6.2. Ejemplos de una distribución beta puntiaguda (a) y aplanada (b). . | 122 |
| 6.3. Primer experimento comparando los modelos EM-MAP y MCMC para realizar la transferencia de información en el reconocimiento de actividades. | 127 |
| 6.4. Segundo experimento comparando los modelos EM-MAP y MCMC para realizar la transferencia de información en el reconocimiento de actividades. | 128 |
| 6.5. Rendimiento de la propuesta para realizar la transferencia de información en la detección de anomalías. | 133 |
| A.1. Esquema de cadenas de Markov y Modelos Ocultos de Markov (HMM). | 172 |

Índice de tablas

| | |
|---|----|
| 2.1. Resumen de trabajos sobre reconocimiento de actividad. | 14 |
| 2.2. Resumen de trabajos sobre detección de patrones de actividad y anomalías. | 16 |
| 3.1. Configuración de las viviendas monitorizadas. | 31 |
| 3.2. Distribución de sensores por vivienda. | 33 |
| 4.1. Tipos de actividades y porcentaje de instancias por clase para cada uno de los conjuntos de datos. | 57 |
| 4.2. Matriz de confusión que muestra los verdadores positivos (VP), el total de las clases reales (TR) y el total de las clases predichas (TP) para cada clase. | 61 |
| 4.3. Resultados experimentales para el conjunto de datos ‘KasterenA’. Se muestra el valor de <i>F-Measure</i> promediado tras la validación cruzada (expresado en %). | 68 |
| 4.4. Resultados experimentales para el conjunto de datos ‘KasterenB’. Se muestra el valor de <i>F-Measure</i> promediado tras la validación cruzada (expresado en %). | 68 |
| 4.5. Resultados experimentales para el conjunto de datos ‘KasterenC’. Se muestra el valor de <i>F-Measure</i> promediado tras la validación cruzada (expresado en %). | 68 |
| 4.6. Resultados experimentales para el conjunto de datos ‘OrdonezA’. Se muestra el valor de <i>F-Measure</i> promediado tras la validación cruzada (expresado en %). | 69 |
| 4.7. Resultados experimentales para el conjunto de datos ‘OrdonezC’. Se muestra el valor de <i>F-Measure</i> promediado tras la validación cruzada (expresado en %). | 69 |
| 4.8. Matriz de confusión del HMM entrenado mediante EM para el conjunto de datos ‘KasterenA’ empleando la representación <i>Las-tSensor</i> | 71 |

| | |
|--|-----|
| 4.9. Matriz de confusión del modelo SVM para el conjunto de datos ‘KasterenA’ empleando la representación <i>LastSensor</i> | 72 |
| 4.10. Matriz de confusión del modelo híbrido HMM/SVM para el conjunto de datos ‘KasterenA’ empleando la representación <i>LastSensor</i> | 72 |
| 4.11. Resumen de las probabilidades empleadas en el HMM, definidas por el conjunto de parámetros $\theta = \{\pi, A, B\}$, y sus correspondientes distribuciones a priori, definidas por los hiperparámetros $\Psi = \{\eta, \rho, \omega, \nu\}$ | 79 |
| 4.12. Comparativa entre el HMM entrenado mediante EM y el HMM entrenado mediante MCMC. Se muestra el valor de <i>F-Measure</i> promediado tras la validación cruzada (expresado en %). | 83 |
| 4.13. Matriz de confusión del HMM entrenado mediante MCMC para el conjunto de datos ‘KasterenA’ empleando la representación <i>LastSensor</i> | 84 |
| 4.14. Valores promediados de las diferentes métricas para cada uno de los modelos propuestos (expresados en %). | 86 |
| 5.1. Resumen de las distribuciones empleadas para modelar las características, y sus correspondientes distribuciones a priori. | 94 |
| 5.2. Matriz de confusión | 103 |
| 5.3. Evaluación individual de las probabilidades usando el conjunto de datos ‘OrdenezA’. | 105 |
| 5.4. Evaluación individual de las probabilidades usando el conjunto de datos ‘OrdenezB’. | 105 |
| 5.5. Evaluación individual de las probabilidades usando el conjunto de datos ‘OrdenezC’. | 106 |
| 5.6. Rendimiento para cada conjunto de datos, combinando las métricas. | 109 |
| 5.7. Comparativa entre nuestra propuesta y el sistema presentado por Shin <i>et. al.</i> [216]. | 112 |
| 6.1. Ejemplo de sensores representados mediante el espacio de características original y el meta espacio. | 119 |
| 6.2. Conjuntos de datos empleados en la transferencia de información. | 124 |
| 6.3. Rendimiento del HMM entrenado mediante MCMC usando el espacio de características original y el mapeado. | 125 |
| 6.4. Rendimiento del modelo de detección de anomalías usando el espacio de características original y el mapeado. | 131 |
| B.1. Matriz de meta características para los datos del conjunto ‘KasterenA’. Representa qué meta características (columnas) se corresponden con qué sensores (filas). | 181 |

| | | |
|------|--|-----|
| B.2. | Matriz de meta características para los datos del conjunto ‘KasterenB’. Representa qué meta características (columnas) se corresponden con qué sensores (filas). | 182 |
| B.3. | Matriz de meta características para los datos del conjunto ‘KasterenC’. Representa qué meta características (columnas) se corresponden con qué sensores (filas). | 183 |
| B.4. | Matriz de meta características para los datos del conjunto ‘OrdenezA’. Representa qué meta características (columnas) se corresponden con qué sensores (filas). | 184 |
| B.5. | Matriz de meta características para los datos del conjunto ‘OrdenezB’. Representa qué meta características (columnas) se corresponden con qué sensores (filas). | 185 |
| B.6. | Matriz de meta características para los datos del conjunto ‘OrdenezC’. Representa qué meta características (columnas) se corresponden con qué sensores (filas). | 186 |
| C.1. | Evaluación de la percepción de utilidad del sistema. | 188 |
| C.2. | Evaluación de la percepción de intrusividad del sistema. | 189 |
| C.3. | Evaluación de la ansiedad producida por del sistema. | 189 |
| C.4. | Esfuerzo requerido por el sistema. | 190 |
| C.5. | Intención de volver a usar el sistema. | 190 |

Capítulo 1

Introducción

Nuestra sociedad se encuentra actualmente en medio de un profundo cambio demográfico. Las mejoras y avances que se producen día a día en el ámbito médico están dando como resultado un incremento en la esperanza de vida de la población, por lo que se espera que vayan en aumento el número de problemas derivados de la edad [195]. El número de personas mayores en las sociedades occidentales está incrementándose de forma dramática, estimándose que para el año 2025 mil millones de personas superará los 60 años de edad [180]. Debido a esta situación, hoy día existe un creciente interés en desarrollar mecanismos y tecnologías que fomenten y permitan vivir de forma independiente a las personas mayores durante el mayor tiempo posible. Dicho interés no viene únicamente derivado del coste en términos de gasto público que supone el aumento de edad de la población [269], sino que recientes estudios muestran que el 71 % de las personas mayores desean enérgicamente envejecer en sus hogares, y otro 12 % preferiría permanecer en ellos durante tanto tiempo como fuese posible [15].

Las tecnologías para la monitorización automática de entornos domésticos son consideradas como una prometedora solución para afrontar los problemas derivados del envejecimiento de la población [91], y los sistemas de inteligencia ambiental son actualmente una pieza fundamental de dicho modelo de interacción [164]. Este tipo de sistemas para la monitorización automática de la salud permiten obtener de una atención eficiente y minuciosa, cumpliendo una doble función: (1) aumentar la seguridad y la sensación de protección de las personas que vivan de forma independiente, y (2) permitir a las personas mayores disfrutar de su autonomía durante más tiempo, mejorando su calidad de vida.

De entre los diversos mecanismos que existen para detectar situaciones anómalas en el comportamiento de un individuo, hay dos que están actualmente especialmente presentes en la literatura: el modelado de Actividades de la Vida Diaria (AVD)[133, 6] y la detección de patrones atípicos [39].

Las AVD es un término ampliamente usado en el ámbito académico para hacer

referencia a un grupo de actividades consideradas definitorias del comportamiento general de un individuo. Son actividades que se llevan a cabo con una periodicidad diaria, y dentro de esta definición se incluyen actividades tales como asearse, alimentarse, bañarse, desplazarse, etc. Las AVD son un pilar fundamental de la funcionalidad de las personas, y por ello profesionales médicos creen que uno de los mejores modos para detectar la aparición de enfermedades en sus primeras fases es mediante la identificación de cambios en este tipo de actividades [142]. Actualmente, los profesionales de la salud únicamente pueden hacer un seguimiento de las AVD de un individuo a través de encuestas y estudios observacionales, dando como resultado una medidas altamente subjetivas. Sin embargo, los sistemas de inteligencia ambiental permiten hacer un seguimiento automático de estas actividades, facilitando datos objetivos del estilo de vida de la persona. Los cambios en la frecuencia o duración de las AVD son considerados como importantes indicadores de trastornos físicos o cognitivos [91].

Con el nombre de detección de patrones atípicos se hace referencia a la problemática de encontrar valores sobre una serie de datos que no se ajusten a un comportamiento esperado. Las técnicas que abordan este problema están presentes en una amplia variedad de áreas y dominios de aplicación, tales como la seguridad informática, el control del fraude o los sistemas de salud. El análisis que ofrecen este tipo de métodos supone una relevante fuente de información acerca de los patrones de comportamiento de una persona [259]. Por ejemplo, patrones de sueño irregulares o variaciones anómalas en el uso del baño pueden ser señales significativas de algún tipo de trastorno [226]. Por tanto, tienen un importante valor potencial en los casos donde es recomendable el seguimiento del estado de salud de una persona a largo plazo.

Estos dos conceptos han sido empleados en el ámbito de la minería de datos como herramientas para evaluar el estado de personas mayores y enfermos crónicos, y están profundamente relacionados con los nuevos desarrollos de sistemas de asistencia en el hogar.

En este contexto, existe una creciente literatura que aborda el estudio de las AVD desde la perspectiva del aprendizaje supervisado en general, y desde el reconocimiento de actividades en particular [128]. Esto es, se busca evaluar el bienestar físico y cognitivo de una persona reconociendo las AVD que realiza, gracias a modelos entrenados con información previa de sus actividades.

Por otro lado, los algoritmos de detección de anomalías o valores atípicos también se han mostrado útiles como métodos de aprendizaje del comportamiento, especialmente cuando no existe conocimiento a priori [246]. Es por eso que dichos métodos están comúnmente enmarcados dentro del aprendizaje no supervisado. El poder establecer un conjunto de patrones definitorios del comportamiento de una persona en concreto, para controlar y dar aviso de acciones inadecuadas o peligrosas, es algo que se ha demostrado útil y sobre lo que se está dedicando

mucho esfuerzo [251].

Las propuestas incluidas en esta tesis doctoral tienen como objetivo el estudio y la optimización de modelos probabilísticos que puedan emplearse para aprender comportamientos típicos de una persona, y que a posteriori sean útiles para determinar conductas anómalas o potencialmente dañinas.

Sin embargo, establecer mecanismos de supervisión del comportamiento, habitualmente supone un conflicto con la privacidad de las personas. Los métodos de monitorización más convencionales son aquellos que emplean grabación de video o audio. No obstante, diversos estudios han demostrado que cámaras y micrófonos son los medios que generan un mayor rechazo por los usuarios [253]. Tradicionalmente, en los sistemas para la monitorización automática de la salud, se han empleado sensores de carácter médico y aparatos de seguimiento remoto específicos para esos sensores [191]. La desventaja de emplear dichos dispositivos es que su inherente especificidad conlleva que sean bastante costosos y requieran de un mayor nivel de colaboración, generando así poca aceptación entre las personas que no tienen dolencias de tipo crónico. El enfoque de este trabajo es emplear tecnologías accesibles y asequibles para su público objetivo. La red de monitorización empleada se compone de un número no predefinido de sensores inalámbricos que ofrecen lecturas sencillas, convirtiendo dicha red en una plataforma barata, escalable y fácilmente configurable.

1.1. Objetivos

El objetivo general de esta tesis doctoral es el diseño y aplicación de modelos de aprendizaje automático para aprender y poder predecir los comportamientos habituales de una persona, haciendo uso de entornos instrumentados con sensores, con el propósito de identificar acciones atípicas o inadecuadas.

La finalidad última del trabajo es el desarrollo de una tecnología que permita mejorar la calidad de vida de las personas mayores que vivan de forma independiente en sus hogares, identificando los diversos medios de ayudarles en el desempeño de sus actividades diarias y permitiendo el seguimiento de estas personas a largo plazo.

El trabajo se basa en un dominio bien conocido y definido por la comunidad investigadora, el modelado del comportamiento humano en un entorno instrumentado, para el que existen diversas propuestas enmarcadas dentro del aprendizaje automático. Sobre dicho dominio se plantean cinco cuestiones principales.

¿Qué técnicas de aprendizaje supervisado pueden emplearse para modelar el comportamiento de una persona? Reconocer las actividades que realiza una persona es una propuesta muy común en la literatura, sin embargo, conocer cuáles son

los mejores modelos para reconocer automáticamente las actividades empleando información de sensores es todavía un problema sin resolver. El diseño de modelos de clasificación requiere en general de un compromiso entre complejidad, coste computacional y volumen de datos requeridos.

Partiendo de la consideración de que muchos comportamientos humanos pueden ser descritos de forma precisa usando conjuntos de modelos dinámicos secuenciados mediante una cadena de estados, nuestra propuesta se orienta hacia el uso de modelos probabilísticos que puedan contemplar las características de dichos modelos dinámicos.

Para abordar esta tarea tomamos como punto de partida una métrica tan popular como las AVD, y unos modelos probabilísticos tan bien conocidos y establecidos como los modelos de Markov. Los HMM ya se han demostrado útiles en el reconocimiento de actividades, y el estudio de este trabajo sobre dichos modelos se centra en:

- Mejorar los procesos de inferencia de parámetros para los HMM en el ámbito del reconocimiento de actividades.
- Presentar mejoras sobre los HMM clásicos, basándose en las propuestas de algoritmos híbridos.

De forma similar a la primera cuestión, nuestra siguiente pregunta es: *¿Qué técnicas de aprendizaje no supervisado pueden emplearse para modelar el comportamiento de una persona?*. Recientemente, se han presentado diferentes sistemas cuyo objetivo es modelar los patrones de las personas mediante diversas métricas relacionadas con los entornos instrumentados. La base de estas propuestas típicamente son los algoritmos de detección de anomalías. Nuestra propuesta es enfocar el estudio del comportamiento desde el marco del aprendizaje no supervisado, presentando un algoritmo capaz de detectar automáticamente patrones de actividad, sin necesidad de haber definido de forma previa ejemplos o instancias de cada actividad. Para ello es necesario determinar qué parámetros son adecuados para estimar patrones de comportamientos habituales y detectar desviaciones de forma precisa.

A su vez, estos planteamientos conllevan una tercera cuestión: *¿Cómo podemos evaluar el rendimiento de estas técnicas para asegurarnos que son suficientemente precisas en el mundo real?* Es muy complicado capturar la estructura y dinamicidad de las actividades y comportamientos humanos dentro de un entorno simulado. Obtener datos auténticos, generados a través de un entorno sensado real, es la mejor solución en este caso, sin embargo, dichos datos son caros y difíciles de conseguir de un modo preciso.

En este trabajo se ha diseñado e implementado un entorno instrumentado propio para aumentar la capacidad sensorial de cualquier localización y obtener información sobre actividades humanas directamente en los hogares de las personas. El sistema consta de sensores lógicos no invasivos y puede ser fácilmente instalado en cualquier ambiente doméstico. Diversos conjuntos de datos provenientes de comportamientos reales se generarán usando este sistema y se pondrán públicamente disponibles a la comunidad investigadora, proporcionando un buen entorno de pruebas donde comparar resultados.

¿Cómo podemos reusar estos métodos para modelar actividades en múltiples hogares? Este tipo de técnicas se enfocan a modelar las características de una única persona. Las diferencias entre el diseño de las viviendas o entre los propios comportamientos de las personas evitan que el modelo aprendido en un entorno pueda trasladarse sin problemas a otro.

En este trabajo se presenta la aplicación de un algoritmo de transferencia de la información, basado en los hiperparámetros de los modelos probabilísticos, para poder transferir el conocimiento obtenido de un hogar a otro, en nuestros dos problemas. Se busca definir hiperparámetros genéricos, que puedan ser comunes a los modelos de reconocimiento de actividad y de detección de patrones.

1.2. Organización de la Memoria

Con la finalidad de dar una visión general de los temas que se tratan en esta tesis doctoral, en el capítulo 2 se profundiza en los modelos de reconocimiento de actividades y detección de patrones, presentando el estado del arte de estos sistemas, sus variantes y características más relevantes.

En el capítulo 3 se presentan los sensores y arquitectura de nuestro sistema de monitorización y el método de anotación empleado para etiquetar nuestros conjuntos de datos.

En el capítulo 4 se presenta nuestra propuesta para el reconocimiento de actividades mediante el uso de modelos híbridos, los cuales tienen propiedades generativas y discriminativas. También se presenta un método de aprendizaje basado en inferencia bayesiana orientado a la aplicación de los HMM en nuestro dominio.

En el capítulo 5 se presenta nuestra propuesta para la detección automática de anomalías en el comportamiento de una persona, mediante el uso de varias métricas probabilísticas y un método estadístico no supervisado encargado de la identificación de valores atípicos.

En el capítulo 6 se detalla nuestra propuesta para aplicar los métodos de modelado del comportamiento presentados en múltiples entornos monitorizados mediante el uso de la transferencia de información. El objetivo de esta propuesta es

poder aplicar el conocimiento previo adquirido en una tarea, para mejorar tareas relacionadas.

Finalmente, en el capítulo 7 se detallan las conclusiones generales, las aportaciones de esta tesis doctoral y las líneas de trabajo futuras.

Capítulo 2

Estado del Arte

Los últimos avances y tendencias tecnológicas están permitiendo que las computadoras abandonen su tradicional rol de electrodomestico de escritorio para pasar a integrarse en diversos aspectos de nuestra vida diaria. Hay dos factores que están facilitando dicha integración: (1) la progresiva reducción del tamaño de los dispositivos electrónicos y (2) el abaratamiento, y por tanto mayor accesibilidad, a todo tipo de tecnologías de sensores.

Estas recientes mejoras tecnológicas están abriendo nuevas áreas de investigación en los sistemas de monitorización y en los entornos tecnológicos instrumentados. De entre las nuevas aplicaciones que están surgiendo en este dominio, destacan las relacionadas con el ámbito de la salud y, en concreto, la asistencia de personas mayores en el hogar [167]. Dichas aplicaciones se basan principalmente en los diferentes tipos de información contextual que se puede obtener del entorno, tales como localización, identidad o acción del usuario [1].

Actualmente, los mecanismos más comunes para la asistencia remota a personas mayores en el hogar son sistemas de alarmas basados en un botón o cordón de pánico. Estos sistemas están ampliamente implantados y existen diversas propuestas comerciales [167]. No obstante, su utilidad depende en gran medida de la capacidad del usuario para reconocer una emergencia, y requiere que el propio usuario active la alarma.

Sin embargo, esta tesis doctoral se enfoca al desarrollo de un sistema de asistencia más acorde al concepto de inteligencia ambiental, es decir, a aumentar la capacidad computacional y sensorial de una localización, para obtener el procesamiento automático de la información contextual. Tomando como base el reconocimiento de actividades y la detección de patrones, se desarrolla un sistema de asistencia ubicuo y autónomo, que permite el seguimiento de personas mayores a largo plazo, gracias a la caracterización y modelado de comportamientos considerados saludables.

El objetivo de este capítulo es brindar una visión general del área en la que

se enmarca este trabajo. A manera de introducción, en la sección 2.1 se detallan los diferentes tipos de tecnologías que existen para la creación de entornos instrumentados. Posteriormente, en la sección 2.2 se presentan diversas propuestas de algoritmos de aprendizaje supervisado relacionados con el reconocimiento de actividades. La sección 2.3 se enfoca a la variante no supervisada vinculada con el modelado de actividades, los sistemas de detección de patrones. En la sección 2.4, se muestran las propuestas de transferencia de información que existen para realizar el aprendizaje de actividades a mayor escala. En la sección 2.5 se enumeran algunos ejemplos de proyectos de entornos inteligentes y de conjuntos de datos públicos, disponibles para la comunidad investigadora. Por último, en la sección 2.6 se muestran algunos ejemplos de sistemas de monitorización comerciales que se encuentran disponibles actualmente en el mercado.

2.1. Tecnologías para la Monitorización Ubicua

Uno de los primeros aspectos a considerar en el diseño de un sistema ubicuo es el tipo de sensores que se van a emplear. Los dispositivos que se usen para la captura de datos definen el nivel más bajo del sistema, y su elegibilidad se determina en base a tres criterios:

Primero, debido a que se monitoriza el espacio personal de una persona, es importante evaluar cómo de intrusivo es percibido el sistema [213]. Un entorno de monitorización puede ser rechazado por un usuario, por ejemplo, debido a que los sensores sean visibles, a la forma de interactuar con éstos o a que la información que capturan sea considerada demasiado privada.

Segundo, la facilidad para desplegar el sistema es relevante. Los sensores que son fáciles de instalar aumentan la aceptación por parte de los usuarios y reducen el coste global del sistema.

Tercero, el tipo de información que capturan los sensores es muy importante. En función de cómo de informativos sean los datos obtenidos, se podrá llevar a cabo el reconocimiento de actividades y la detección de patrones con mayor o menor éxito.

En esta sección se profundiza en los tipos de sensores más empleados en la literatura, detallando cómo se adaptan a los criterios mencionados.

2.1.1. Dispositivos Portátiles

Este tipo de tecnología, originalmente conocida como *Wearable Computing*, hace referencia a una serie de dispositivos que el propio usuario puede llevar encima, integrados en su espacio personal, controlados por él mismo y que están en constante interacción [156].

Esta clase de sensores son, por lo general, bastante compactos y tienen funcionalidades enfocadas principalmente a la captura y envío de datos. El tipo de dispositivo llevable más común en la literatura es el giróscopo, ya que permite medir la dirección y velocidad de un movimiento concreto del usuario [173, 178, 264]. Otros sensores también presentes en diversos estudios son los sensores de luz, de humedad, cardíacos y brújulas [189, 147]. Este tipo de tecnología ha resultado especialmente útil en proyectos donde el objetivo es monitorizar señales biomédicas muy específicas [183], gracias a la combinación de prendas y dispositivos integrados, o lo que ha venido a denominarse como “ropa biomédica”. También existen trabajos donde se emplea un enfoque distribuido, instalando un sensor en diferentes miembros del paciente, para así obtener medidas más diversas y específicas [152].

Un inconveniente al que se enfrenta esta tecnología es su inherente intrusividad, puesto que para emplear un sistema de estas características es muy importante mantener un correcto equilibrio entre la comodidad del usuario y la usabilidad de los sensores. Un estudio sobre la aceptación de este tipo de dispositivos estableció que los lugares donde los usuarios mejor toleran llevar un sensor son la cadera y el pecho [179]. Sin embargo, debido a la reciente proliferación de teléfonos inteligentes o *smartphones*, dicha situación está empezando a cambiar y se están abriendo nuevos campos para la investigación con esta tecnología. Los diferentes tipos de sensores integrados en los teléfonos inteligentes, tales como acelerómetros, sistemas de posicionamiento global, grabadoras de audio o videocámaras, junto con su pequeño tamaño y capacidad de cómputo, convierten estos dispositivos móviles en buenas herramientas para la monitorización ubicua [140].

Los datos recogidos por estos dispositivos principalmente ofrecen información sobre la postura y movimiento del usuario. Debido a la limitada información sobre la ubicación de la persona o los objetos con los que interacciona, es bastante complicado llevar a cabo reconocimiento de actividades basándose únicamente en estos datos. Actividades con movimientos muy definidos como pasar la aspiradora o beber de un vaso sí pueden modelarse [189]; sin embargo, para reconocer otras actividades como correr o andar es necesario realizar el aprendizaje con la premisa de conocer la localización de los sensores sobre el cuerpo del usuario [161].

La información que ofrecen estos sensores también puede procesarse como una serie temporal multivariada y emplearse para la detección de patrones. En un trabajo llevado a cabo por Thad et al., este tipo de sistema se empleó con éxito para detectar automáticamente patrones de actividad haciendo uso de un modelo probabilístico entrenado mediante el algoritmo EM (Esperanza-Maximización) [165].

2.1.2. Videocámaras

Las videocámaras son instrumentos muy populares dentro del ámbito de la computación ubicua, dado que permiten obtener datos muy informativos de alta dimensionalidad de un modo relativamente simple [168]. Son dispositivos ampliamente utilizados para el reconocimiento de actividades, existiendo diversos estudios donde se describe su utilidad para reconocer actividades y acciones primitivas, las cuales no son susceptibles de ser descompuestas [193, 239].

A partir de las imágenes capturadas por una videocámara puede extraerse una valiosa información, por ejemplo la localización o trayectoria de un individuo, que sirva como base para llevar a cabo tareas de modelado de comportamiento. Existen en la literatura diversos ejemplos de cómo emplear la ubicación de un usuario percibida por una cámara para determinar la actividad que éste está realizando [48] o detectar un patrón de comportamiento [155]. En un estudio llevado a cabo por Duong et al. la imagen es dividida en cuadrículas y un sistema multicámara es capaz de detectar qué celdas se ven afectadas por el movimiento del usuario. Celdas significativas están relacionadas con objetos de interés y son útiles para determinar qué actividad está realizando el usuario [70]. Un sistema multicámara es también empleado en diversos estudios para determinar patrones de comportamiento, ya sea modelando enlaces temporales entre los puntos de entrada y salida de las imágenes para aprender trayectorias normales y detectar desviaciones anómalas de las mismas [72, 73], o mediante localización, determinando áreas de interés en la imagen y analizándolas para detectar periodos de inactividad [172].

Además de la localización, existen otros indicadores que son empleados para el reconocimiento de actividad usando videocámaras, por como ejemplo el objeto que la persona esté manipulando. Normalmente una sola cámara se centra en un punto particular del entorno, como por ejemplo el lavabo o la vitrocerámica, en el caso del hogar, y los objetos utilizados se emplean para reconocer actividades como hacer té, tomar una medicina [258] o beber de un vaso [163].

2.1.3. Identificación por Radiofrecuencia

La tecnología de identificación por radiofrecuencia, o RFID (*Radio Frequency IDentification*), es una tecnología para detectar y recuperar información remotamente de dispositivos conocidos como etiquetas o *tags* RFID.

Existen dos tipos de etiquetas RFID, pasivas o activas. Las etiquetas pasivas obtienen energía de la señal de radio frecuencia emitida por un lector RFID y usan dicha energía para devolver un flujo de datos. Las etiquetas activas, por el contrario, están equipadas con una fuente de energía propia (generalmente una batería), haciendo posible la lectura de los datos a distancias mayores [63].

El fabricante de circuitos integrados Intel Corporation ha empleado este tipo

de tecnología para desarrollar un sistema de monitorización enfocado al reconocimiento de actividades. El sistema se compone un lector RFID diseñado en forma de pulsera, denominado *iBracelet*, que se puede llevar en una o ambas muñecas [141]. Los objetos considerados de interés son etiquetados usando etiquetas pasivas, y gracias a ellas el *iBracelet* permite identificar qué objetos son manipulados por el usuario. Debido a la información que contienen las etiquetas acerca de los objetos, este sistema permite desplegar una red de monitorización de forma barata, permitiendo además detectar acciones primitivas de modo bastante básico, para realizar reconocimiento de actividades [34].

Un inconveniente de este sistema es su intervalo de lectura, ya que los errores en la lectura de etiquetas son comunes. En una comparación entre diferentes sistemas de monitorización, esta tecnología obtuvo los peores resultados debido a esta limitación [151]. Estos sistemas también aparecen en la literatura en combinación con otro tipo de sensores, como los acelerómetros. Midiendo a la vez el tipo de objeto y movimiento, es posible compensar las deficiencias del sistema, obteniendo un resultado mejor en el reconocimiento de actividades [220].

2.1.4. Redes de Sensores Inalámbricos

Una red de sensores inalámbricos (WSN, por sus siglas en inglés) consiste en una colección de dispositivos de monitorización autónomos y espacialmente distribuidos. Cada dispositivo o nodo que componga la red debe tener la capacidad de obtener, procesar y transmitir información [217]. Debido a su flexibilidad, bajo coste y fácil despliegue, actualmente esta tecnología es muy popular en dominios donde se necesita algún tipo de monitorización ubicua. Por ejemplo, este tipo de sistemas se han aplicado con éxito en áreas tan diferentes como la telemedicina [93], la ecología del comportamiento [92] o la agricultura supervisada [166].

En estos sistemas la comunicación entre nodos no se ve afectada por la latencia y requiere de poco ancho de banda, por lo que se pueden emplear protocolos de comunicación muy eficientes energéticamente [240]. Los protocolos de enrutamiento de estas redes permiten a los nodos configurarse de forma dinámica, sin necesidad de definir topologías previamente [31].

Los nodos de estas redes pueden trabajar con una gran variedad de sensores diferentes. En un entorno doméstico, es común emplear distintos tipos de sensores para la medición de diferentes eventos: sensores infrarrojos de movimiento para control de áreas, sensores magnéticos para control de puertas, sensores de nivel para control de flujos, sensores de presión para control de áreas de descanso o sensores de corriente eléctrica para control del uso de electrodomésticos [127, 256].

Las redes de sensores inalámbricos ofrecen varias ventajas con respecto al resto de tecnologías para la supervisión del hogar. Primero, los sensores pueden

instalarse de forma oculta al usuario, limitando la intrusividad del sistema. Segundo, la información almacenada por estos sensores es inherentemente anónima y la cantidad de datos sensibles almacenados es muy limitada. Tercero, el despliegue de la red se realiza de forma rápida y sencilla, sin necesidad de fuentes de energía o preinstalaciones de ningún tipo. Estas valiosas cualidades convierten a este tipo de redes en una de las tecnologías más prometedoras para establecer mecanismos de supervisión en un entorno doméstico [53]. De hecho, este dominio está considerado como uno de los mercados de más rápido crecimiento para las tecnologías basadas en redes inalámbricas [119].

2.2. Reconocimiento de Actividad

El reconocimiento de actividades se enmarca dentro de la minería de datos como una tarea de clasificación supervisada. Partiendo de un conjunto de actividades conocidas previamente, se definen y extraen las características relevantes de la información disponible, para generar un modelo que permita clasificar nuevos datos de ejemplo. Respecto al conjunto de actividades, las AVD son las que han adquirido una mayor popularidad por parte de la comunidad investigadora debido a su utilidad para medir la capacidad funcional de las personas [34, 220, 127, 70, 163].

Uno de los primeros trabajos relevantes en el área enfocaba esta tarea como un problema de reconocimiento de planes, aplicando modelos lógicos [134]. Kautz formalizó el problema, usando el método de razonamiento no-monótono denominado circunscripción propuesto por McCarthy [162], con el propósito de explicar cada secuencia de observaciones con el mínimo número de planes. Sin embargo, en dicho estudio se asume un conjunto de datos sin ruido ni datos irracionales, lo cual no se corresponde con la realidad. Para abordar el problema de un modo más apropiado, Bouchard *et. al.* enfocan el problema hacia el reconocimiento de actividades en personas con discapacidades cognitivas [25]. Usando teoría de conjuntos y descripción lógica de acciones se generan dinámicamente nuevos planes de acciones estadísticamente probables, para así intentar identificar acciones incoherentes.

Chen *et. al.* [45] también emplean métodos lógicos para reconocer actividades, en este caso mediante cálculo de eventos. Definen un evento como la observación de una acción a través de los sensores, un predicado como la relación entre entidades del mundo y una actividad como la combinación de varios predicados. El perfil del usuario sirve para definir un comportamiento normal y, si el sistema detecta alguna desviación sobre dicho comportamiento, emite una señal de aviso.

Otra técnica que ha ofrecido buenos resultados en el área, y se ha convertido en un enfoque especialmente popular, es el uso de modelos probabilísticos. Uno

de los primeros trabajos que empleó este tipo de modelos con sistemas de monitorización automática lo propusieron Tapia *et. al.* [230]. En dicho estudio emplearon algoritmos bayesianos para modelar cada sensor de forma independiente, sin incluir ningún tipo de información sobre la secuencialidad de las lecturas [151]. Barnan *et. al.* usaron una colección de algoritmos de aprendizaje bien conocidos, tales como árboles de decisión y reglas de clasificación por vecindad, para clasificar un conjunto de acciones primitivas que fueron definidas manualmente, de forma previa, por varios expertos médicos [62].

No obstante, el reconocimiento de las actividades humanas posee ciertas características que lo convierten en un problema no trivial: cada usuario realiza cada actividad específica de un modo personal y diferente, la duración de las actividades es habitualmente desconocida y los datos generados por los sensores son ruidosos y tienen un comportamiento no determinista.

Existen propuestas enfocadas específicamente a hacer frente a alguno de estos retos. Iglesias *et. al.* postulan que, debido a que las personas llevan a cabo las actividades de un modo dinámico, los clasificadores evolutivos basados en reglas difusas [3] son adecuados para abordar dicho problema y modelar las AVD, segmentando la información en secuencias y almacenándola en estructuras de datos tipo árbol [113].

Sin embargo, los modelos probabilísticos temporales han demostrado ofrecer un buen comportamiento y ser capaces de manejar mejor la incertidumbre causada por las características de este dominio [132]. Estos modelos son capaces de tomar una decisión informada acerca de qué actividad está realizando el usuario basándose en la evidencia que los sensores proporcionan. Un modelo oculto de Márkov o HMM (por sus siglas en inglés, *Hidden Markov Model*) es un modelo probabilístico temporal muy popular, ampliamente presente en la literatura, donde cada estado se describe por una o más variables aleatorias discretas. Los HMM han sido aplicados con éxito en muchos problemas de modelado de datos secuenciales, tales como el reconocimiento de voz [203], reconocimiento de escritura [10] y análisis de secuencias biológicas [254].

Patterson *et. al.* emplearon HMM para el reconocimiento de actividades usando *tags* RFID y probaron cómo incluyendo las correlaciones temporales en el aprendizaje se obtiene una mejora en la precisión del modelo [185]. Wilson *et. al.* llevaron a cabo el reconocimiento de actividades en un solo entorno doméstico donde habitaban varias personas. Su propuesta es similar al trabajo de Patterson, en el sentido en que el HMM se usó para modelar correlaciones temporales; sin embargo, también empleó un filtro de partículas para asignar cada lectura de cada sensor a un usuario específico. Después, se empleó el modelo para reconocer las actividades de cada persona individualmente [252].

Recientemente, Kasteren *et. al.* han empleado HMM con redes de sensores inalámbricos para modelar las AVD de personas mayores que viven de forma

| Artículo | Descripción | Modelo | Dominio/Instrumentación |
|--------------------------------------|---------------------------------|--|-------------------------|
| [261] | Reconocimiento de actividad | Modelo Oculto de Markov | Video |
| [44] | Reconocimiento de actividad | Máquina de vectores de soporte | Video, audio |
| [201] | Reconocimiento de actividad | Conjuntos de clasificadores | Acelerómetros |
| [154] | Reconocimiento de actividad | Árboles de decisión Aprendizaje de instancias | Acelerómetros |
| [203] [206] [80] [237] [169] [84] | Reconocimiento del habla | Modelo híbrido HMM/RNA | Audio |
| [108] [221] | Reconocimiento del habla | Modelo híbrido HMM/Redes Neuronales Recurrentes | Audio |
| [202] | Reconocimiento del habla | Modelo Oculto de Markov Jerárquico | Audio |
| [212] | Predicción de series temporales | Modelo híbrido HMM/RNA | Láseres |
| [75] [159] | Reconocimiento de caracteres | Modelo híbrido HMM/RNA | Texto |
| [144] [159] | Modelado de señas biomédicas | Modelo híbrido HMM/RNA | Sensores biomédicos |
| [82] | Reconocimiento de dígitos | Modelo híbrido HMM/SVM | Texto |
| [218] [198] | Reconocimiento del habla | Modelo híbrido HMM/SVM | Audio |
| [85] | Predicción de series temporales | Modelo híbrido HMM/SVM | Sensores biomédicos |
| [158] [157] | Reconocimiento del habla | Modelo híbrido HMM/Redes bayesianas | Audio |
| [147] | Reconocimiento de actividad | Modelo híbrido HMM/RNA | Red de sensores |
| [220] | Reconocimiento de actividad | HMM, Modelos bayesianos | Acelerómetros, RFID |
| [34] | Reconocimiento de actividad | Modelo Oculto de Markov | RFID |
| [127] [128] | Reconocimiento de actividad | Modelo Oculto de Markov Campo aleatorio condicional | RFID |
| [163] | Reconocimiento de actividad | Modelos probabilísticos | Video |
| [70] | Reconocimiento de actividad | Modelo Oculto de Semi-Markov | Video |
| [230] | Reconocimiento de actividad | Algoritmos bayesianos | Red de sensores |
| [62] | Reconocimiento de actividad | Árboles de decisión Reglas de decisión | Red de sensores |
| [113] | Reconocimiento de actividad | Modelos evolutivos | Red de sensores |
| [203] | Reconocimiento del habla | Modelo Oculto de Markov | Audio |
| [10] | Reconocimiento de caracteres | Modelo Oculto de Markov | Texto |
| [185] [252] | Reconocimiento de actividad | Modelo Oculto de Markov | RFID |
| [11] | Reconocimiento de actividad | Aprendizaje de instancias | Acelerómetros |
| [160] | Reconocimiento de actividad | Árboles de decisión | Acelerómetros |

Tabla 2.1: Resumen de trabajos sobre reconocimiento de actividad.

independiente [128]. Discretizando las lecturas obtenidas de los sensores en intervalos de tiempo de longitud definida, obtienen un vector de datos para cada instante. Entrenando modelos temporales probabilísticos mediante el algoritmo de *Baum-Welch* [197], establecen un patrón de correspondencia entre cada vector de lecturas con una de las posibles actividades.

Además de los modelos temporales probabilísticos clásicos, también se han presentado estudios basados en modelos híbridos para reconocer actividades mediante el uso de dispositivos llevables. Los resultados experimentales mostraron una mejora en la precisión del modelo híbrido con respecto a los modelos más tradicionales [147].

Otro tipo de algoritmos de aprendizaje bien conocidos, como por ejemplo los basados en instancias, también se han empleado para este problema. Uno de los métodos basados en instancias más conocidos es el *vecino más cercano* (*Nearest*

Neighbor o NN en inglés) [56], en donde cada nueva secuencia de observaciones se compara con las observadas previamente y se clasifica según la secuencia más similar. Este procedimiento tan simple ha dado a menudo buenos resultados. Bao *et. al.* compararon este método con otros tipos de algoritmos de clasificación para reconocer actividades empleando datos obtenidos a través de acelerómetros. Concluyeron que el método NN obtiene peores resultados que los árboles de decisión, otro método de clasificación bien establecido que divide los datos de entrenamiento estableciendo reglas basándose en su clase [11]. Maurer *et. al.* [160] también emplearon árboles de decisión para aprender descripciones lógicas de un conjunto de actividades tomando como entradas los datos de un sensor que los usuarios llevaban en la muñeca. Los árboles de decisión ofrecen la ventaja de generar modelos fácilmente interpretables, pero su rendimiento no es bueno cuando el espacio de características lo definen datos numéricos de alta precisión.

Para dar una visión general sobre el área, en la Tabla 2.1 se ofrece un resumen de los trabajos relativos al reconocimiento de actividad incluidos en la presente tesis doctoral.

2.3. Detección de Patrones de Actividad

Los algoritmos de reconocimiento de patrones son métodos de aprendizaje no supervisado cuya finalidad es encontrar automáticamente patrones recurrentes en un conjunto de datos secuenciales. Una importante ventaja de este tipo de algoritmos es que son capaces de trabajar buscando pautas en conjuntos de datos sin etiquetar.

Una aplicación muy común de estos algoritmos de detección es la búsqueda de patrones en conjuntos de datos con ruido. Es posible, por ejemplo, la detección automática de series de palabras recurrentes en una señal de audio [177] o emplearlos para integrar datos de expresión génica [28]. La señal continua es típicamente procesada a priori, transformando los datos y obteniendo una representación simbólica discreta de los mismos. De este modo se evita tener que realizar una búsqueda exhaustiva ya que habitualmente no existe ningún conocimiento a priori sobre los patrones a identificar [47].

Estos algoritmos también son útiles para encontrar correlaciones entre eventos dentro de grandes conjuntos de datos. Cuando se trabaja con entornos tecnológicos instrumentados, la identificación de sensores que demuestren un comportamiento similar durante los mismos intervalos de tiempo, es relevante para llevar a cabo la detección de patrones. Magnusson *et. al.* emplean esta idea para desarrollar una técnica denominada *T-patterns* [153]. El objetivo de esta técnica es calcular la frecuencia promedio de cada uno de los eventos presentes en un conjunto de datos, asumiendo que dichos eventos son independientes y están aleatoriamente

| Artículo | Descripción | Tipo de aprendizaje | Método |
|-------------------|---|---------------------|---|
| [81] | Modelado del comportamiento normal | Semisupervisado | Métodos kernel |
| [232] | Generalización sobre un conjunto de entrenamiento | Supervisado | Redes de neuronas |
| [255] [120] | Generalización sobre un conjunto de entrenamiento | Supervisado | Redes bayesianas |
| [143] [30] | Generalización sobre un conjunto de entrenamiento | Supervisado | Máquinas de soporte de vectores |
| [227] | Generalización sobre un conjunto de entrenamiento | Supervisado | Modelos estadísticos paramétricos |
| [266] [4] [112] | Generalización sobre un conjunto de entrenamiento | Supervisado | Técnica del vecino más cercano |
| [244] | Agrupamiento | Semisupervisado | Ordenación mediante ranking |
| [33] | Agrupamiento | No supervisado | Búsqueda de subsecuencia común más larga |
| [5] [12] | Modelado del comportamiento normal | Supervisado | Técnicas estadísticas gaussianas |
| [43] | Modelado del comportamiento normal | Supervisado | Técnicas estadísticas de regresión lineal |
| [2] | Modelado del comportamiento normal | Supervisado | Técnicas estadísticas de mixturas de distribuciones |
| [9] | Modelado de patrones | Supervisado | Estimación de desviaciones |
| [216] [263] [181] | Modelado de patrones | Supervisado | Máquinas de soporte de vectores |
| [246] [247] [90] | Modelado de patrones | Supervisado | Aprendizaje de los ritmos circadianos |
| [74] [107] | Modelado del comportamiento normal | Supervisado | Técnicas estadísticas bayesianas |
| [177] [255] | Modelado de patrones | No supervisado | Técnicas estadísticas bayesianas |
| [28] | Modelado de patrones | No supervisado | Agrupamiento de microarrays |
| [47] [69] | Modelado de patrones | Supervisado | Técnicas estadísticas probabilísticas |
| [153] | Modelado de patrones | Supervisado | <i>T-patterns</i> |
| [187] | Generalización sobre un conjunto de entrenamiento | Supervisado | Conjuntos de clasificadores |
| [268] | Modelado del comportamiento normal | No supervisado | Modelos tipo árbol |
| [149] | Modelado del comportamiento normal | Supervisado | Técnicas heurísticas |

Tabla 2.2: Resumen de trabajos sobre detección de patrones de actividad y anomalías.

te distribuidos en el tiempo. Se considera que dos eventos conforman un patrón cuando la aparición de uno de los eventos condiciona con una alta probabilidad la aparición del otro evento. Cada nuevo patrón detectado, es añadido como un nuevo evento a su vez, permitiendo la aparición de patrones jerárquicos. En un estudio desarrollado por Tavenard *et. al.*, los *T-patterns* fueron aplicados a la detección automática de patrones de movimiento dentro de entornos de inteligencia ambiental [231].

En la literatura existen diversos métodos que tienen la detección de patrones como base de uno u otro modo. La minería de reglas de asociación se basa en el mismo principio, y a los patrones frecuentes los denomina formas primitivas [60, 109]. Diversos algoritmos de identificación y clasificación de series temporales trabajan definiendo prototipos de cada clase, sobre los cuales después cotejar las nuevas instancias [135, 101]. En robótica, Oates *et. al.* presentaron un método para permitir a varios agentes autónomos generalizar patrones desde un conjunto de, lo que ellos denominaron, “experiencias cualitativamente diferentes” [176].

Estos métodos sirven como base para poder detectar desviaciones en el comportamiento de las personas, mediante la definición de patrones normales y el cálculo de anomalías sobre los mismos. La detección de valores atípicos, o de-

tección de desviaciones [139], tiene como finalidad encontrar en un conjunto de datos observaciones o patrones que no se ajusten a unas pautas definidas, o sean significativamente distantes del resto de los datos [39]. Diferentes sistemas de detección de anomalías se han aplicado a multitud de dominios. El caso de la detección de intrusos en sistemas informáticos es un ejemplo recurrente. Uno de los retos para los sistemas aplicados a este dominio es la eficiencia, debido a los grandes volúmenes de datos con los que tiene que tratar, por lo que las propuestas semisupervisadas y no supervisadas son las más abundantes en este caso [88, 105].

También se han presentado sistemas de detección de anomalías orientados a la identificación de fraudes. Existen sistemas en explotación dedicados a la búsqueda de acciones fraudulentas en el uso de tarjetas de crédito [30], teléfonos móviles [187] o en sistemas de inversión [68]. Un inconveniente que presenta este dominio es que debido a la naturaleza del problema, las anomalías no son estáticas, sino que evolucionan con el tiempo del mismo modo que cambia la forma de actuar de los delincuentes.

Un dominio sobre el que se ha despertado en los últimos años bastante interés son las redes de sensores [268, 112, 120, 241, 69]. Las anomalías que se detectan en una red de sensores pueden significar que existe un problema con alguno de los nodos de la red o que se está identificando un evento interesante para ser analizado. El reto que supone este dominio es la necesidad de trabajar con entornos en tiempo real, por lo que los algoritmos de detección deben estar muy orientados hacia la eficiencia.

También ha despertado mucho interés la aplicación de estas técnicas al ámbito de la salud. Tradicionalmente la detección de valores atípicos ha estado relacionado con el análisis de los registros de los pacientes. La mayoría de los trabajos en este área están destinados a detectar valores extraños en los registros, mediante el uso de datos pertenecientes a pacientes sanos, para llevar a cabo el entrenamiento de los modelos [149, 255].

La detección de patrones también se ha empleado para modelar las actividades humanas y realizar análisis del comportamiento [165, 260]. Recientemente, se ha objetivado y actividad [246, 90]. Suzuki *et. al.* proponen identificar las actividades diarias de una persona mayor, calculando las desviaciones típicas sobre las lecturas de un conjunto de sensores [223]. Usando sensores infrarrojos, magnéticos, fotoeléctricos y eléctricos, hace una estimación del número de lecturas por sensor que se corresponden a cada intervalo temporal, y sobre dichos valores se calcula la desviación típica.

Suryadevara *et. al.* usan un planteamiento similar en una red de sensores inalámbricos. Haciendo una estimación del uso de cada sensor mediante el cálculo del tiempo total que permanece activo cada uno, emplean esos datos para detectar desviaciones sobre el comportamiento de la persona, haciendo uso de los que ellos denominan “funciones de bienestar” [222]. Glascock *et. al.* diseñaron un sistema

orientado a pacientes crónicos, que aplicaba estadísticas simples de los sensores, tales como, detectar el primer sensor en activarse por la mañana y el último por la noche para determinar el periodo de sueño, o contar el número de veces que el usuario accedía a su medicación [86].

En la Tabla 2.2 se muestra un resumen con los diferentes trabajos relacionados con la detección de patrones y anomalías que han sido consultados en la presente tesis doctoral.

2.4. Transferencia de Información

Un problema de las técnicas de aprendizaje anteriormente descritas, y de muchos de los métodos presentes en la literatura relacionados con el aprendizaje de observaciones o patrones, es el hecho de necesitar siempre un entrenamiento previo de los algoritmos empleando un dominio específico [71, 127, 216]. En el caso que nos ocupa, esto se traduce en la necesidad de obtener una cantidad suficiente de datos a través de los sensores para poder modelar el comportamiento de una única persona en un entorno instrumentado concreto.

Para desarrollar un sistema de modelado del comportamiento en el hogar realmente escalable, éste debe poder aplicarse a todo tipo de entornos, independientemente de las características y dimensiones de la vivienda donde se instale. Para poder emplear los algoritmos presentados en el modelado del comportamiento de diferentes personas, dichos algoritmos necesitarán adaptarse al hogar de cada una de esas personas. Es necesario algún proceso que permita generalizar modelos enfocados a un hogar concreto, a través de conjuntos de datos obtenidos gracias a otro entorno instrumentado diferente. Este tipo de generalización, donde empleando datos obtenidos en una vivienda se entrenan los modelos para un nuevo entorno, se ha obtenido en trabajos previos gracias a la transferencia de información [126].

Dentro de la minería de datos, la transferencia de información se define como el proceso algorítmico por el cual el conocimiento adquirido durante un proceso de aprendizaje se emplea para mejorar el conocimiento previo acerca de procesos o problemas relacionados [243]. A diferencia del problema clásico del aprendizaje automático, donde se busca generalizar comportamientos a partir de ejemplos de un mismo problema, la transferencia de información se enfoca en transferir conocimientos a través de dominios parecidos, pero no necesariamente iguales. El uso conjunto de conceptos relacionados y relevantes con la tarea objetivo puede mejorar el aprendizaje final de la misma.

En función de la relación entre tarea origen y tarea objetivo, y del tipo de datos disponibles para cada una, se diferencian varias clases de transferencia de información. En el dominio tratado aplican dos de esos tipos principalmente: transfe-

rencia inductiva y transferencia de información no supervisada [182]. En la transferencia inductiva, la tarea principal u objetivo es diferente de la tarea secundaria u origen, sin importar a que dominio pertenecen. En este caso es necesario que existan datos etiquetados en la tarea secundaria para poder inducir un modelo predictivo para la tarea principal. En transferencia de información no supervisada, las tareas principal y secundaria también son diferentes aunque estén relacionadas, pero el aprendizaje está enfocado a resolver problemas no supervisados, no existiendo datos etiquetados en ninguna de las dos tareas [250].

El tipo de transferencia de información que se emplea para la tarea de reconocimiento de actividades es por tanto la transferencia inductiva [36]. En una situación típica de aprendizaje para reconocimiento de actividades se dispone de diversos conjuntos de datos con las actividades de los usuarios bien etiquetadas (tarea origen), y otro conjunto de datos sobre el que se desea realizar el reconocimiento de actividades (tarea objetivo) y que consta de pocos, o ningún, dato etiquetado.

Los primeros estudios sobre transferencia inductiva se enfocaron principalmente en aprendizaje multi-tarea, para poder llevar a cabo distintos procesos de aprendizaje implicados en distintas tareas de forma simultánea, en vez de emplear un proceso de aprendizaje por cada tarea por separado [14, 233]. Un ejemplo clásico de este tipo de aplicación es la clasificación temática de noticias usando grupos de noticias como etiquetas, donde el modelo debe determinar a qué grupo pertenece cada documento [57, 199]. En el ejemplo mencionado, se puede considerar un caso concreto donde se debiera clasificar si una noticia pertenece al grupo ‘Noticias aeroespaciales’ o al grupo ‘Noticias de informática’. En dicho caso, si se incluyera en el proceso de aprendizaje información acerca de noticias sobre religión, beisbol o motocicletas, la precisión del modelo mejoraría, debido a que estas otras noticias aportan información adicional acerca de la coocurrencia de las palabras [199].

En un estudio reciente realizado por Cook *et. al.* [50], se ha llevado a cabo una clasificación muy minuciosa de las técnicas de transferencia de información actuales en función de cuatro aspectos, a saber: tipo de sensores empleados, diferencias entre tareas origen y objetivo, tipo de etiquetado de los datos y representación del conocimiento. Estos mismos autores, en el ámbito del reconocimiento de actividad, demostraron que es posible emplear métodos de transferencia de información supervisados y semisupervisados, para generalizar técnicas de aprendizaje a través de diferentes viviendas [52].

Por otro lado, en la transferencia de información no supervisada, habitualmente se transfiere información sobre la representación y los atributos de los datos [36]. Dai *et. al.* usan este tipo de transferencia para mejorar específicamente el rendimiento de un algoritmo de agrupamiento, llevando a cabo el aprendizaje de forma simultánea en la tarea origen y en la tarea objetivo, para así identificar atri-

butos compartidos [59]. En otro caso, Raina *et. al.* proponen la aplicación de la codificación dispersa [145] para modelar y transferir los atributos de alto nivel del conjunto de datos [200].

La forma óptima de realizar transferencia de la información sigue siendo un problema no resuelto. Sin embargo, un método que funciona especialmente bien con los modelos probabilísticos es aprender la probabilidad a priori de los parámetros, usando las tareas secundarias, y emplear ese conocimiento para inferir los parámetros del modelo de la tarea principal [146, 199]. La probabilidad a priori permite obtener una estimación inicial de los parámetros que definirán el modelo de la tarea objetivo. Este método se ha empleado con éxito en problemas donde los datos son modelados mediante distribuciones multinomiales [130] o mediante distribuciones gaussianas [24].

2.5. Conjuntos de Datos y Proyectos Académicos

Para poder evaluar adecuadamente cualquier algoritmo de minería de datos, son indispensables conjuntos de datos apropiados. En lo relativo a entornos de inteligencia ambiental, existen diversos entornos o casas inteligentes construidos por diferentes equipos de investigación a lo largo del mundo, con el fin de realizar experimentos y generar conjuntos de datos. Estos conjuntos de datos idealmente están generados con datos reales y se encuentran totalmente etiquetados, sirviendo así tanto para el reconocimiento de actividades como para otras tareas.

En esta sección, se revisan los proyectos de casas inteligentes de diversos grupos de investigación, se describen los conjuntos de datos pertenecientes a cada una y se describe el sistema de monitorización empleado en cada caso.

2.5.1. Proyectos de Casas Inteligentes

A día de hoy, existen diversos ejemplos de casas inteligentes construidas para fines principalmente académicos. La mayoría de estos proyectos están enfocados a monitorizar personas mayores con trastornos cognitivos, visuales o motores [228, 219, 32]. Típicamente, en estas casas se instalan diferentes sensores, actuadores o monitores biomédicos, y estos dispositivos almacenan información mediante algún tipo de sistema centralizado.

Proyectos como CASAS Smart Home de la Washington State University [51] han puesto de manifiesto la utilidad de estos sistemas para monitorizar y automatizar la asistencia a personas mayores que vivan dentro de un entorno sensado. El proyecto tiene como objetivos generales minimizar el coste de mantenimiento del sistema y maximizar la comodidad de sus habitantes.

El proyecto MavHome de la Universidad de Texas combina en un mismo lugar diversas tecnologías: sistemas multimedia, inteligencia artificial, computación móvil y robótica [61]. El objetivo de este proyecto es crear un entorno que actúe como un agente racional, maximizando el confort de los habitantes a la vez que reduce los costes operacionales. Dicho agente debe ser capaz de detectar y predecir los hábitos de movilidad de los ocupantes de la casa. Para ello, se emplea un algoritmo denominado Lezi, el cual genera un modelo estadístico que calcula la probabilidad de ciertas acciones de los usuarios durante cada intervalo de tiempo [20].

En la University of Florida, el entorno denominado *Gator-Tech* fue construido como un laboratorio de inteligencia ambiental para personas mayores. Dicho entorno posee diversos dispositivos particulares, tales como un buzón de correo que detecta la llegada de correspondencia, una nevera que controla el consumo de alimentos o un espejo del baño que sirve como calendario [104]. Todos estos dispositivos se integran con otros sensores y actuadores, y además están conectados a una plataforma operacional diseñada expresamente para optimizar el confort y la seguridad de una persona mayor [103].

La iniciativa *Aware Home* es un proyecto del Georgia Institute of Technology. Es un entorno interdisciplinario consistente en dos espacios habitables idénticos e independientes monitorizados mediante videocámaras y sensores de presión. El suelo es un tipo de superficie inteligente, diseñada para detectar los pasos de las personas, y en base a eso construir un modelo para predecir los hábitos de los usuarios. Mientras que los sensores se encuentran instalados en el piso y controlan las áreas de desplazamiento, las videocámaras suponen un método de supervisión constante [136]. Para crear los modelos del comportamiento se hace uso de diversas herramientas, tales como los HMM o las redes neuronales. El principal objetivo de este proyecto es permitir a las personas envejecer en sus hogares, no sólo mejorando su calidad de vida sino también su longevidad.

CarerNet es un proyecto desarrollado en el Reino Unido, donde se integran diversos sistemas de telemedicina y monitorización en los domicilios de los usuarios [251]. Este sistema incorpora una red de sensores y un sistema de monitorización inteligente para capturar y tratar información fisiológica de los usuarios, identificar su estilo de vida y modelar el entorno. El sistema de monitorización está basado en sensores inteligentes, dispositivos acoplables al cuerpo, una unidad central de procesado y en los datos médicos del propio usuario.

En Japón, la empresa NTT DoCoMo ha desarrollado un laboratorio multimedia donde modelar e identificar el comportamiento de sus usuarios, empleando una red de sensores y objetos etiquetados mediante tecnología RFID. Las actividades diarias de los usuarios son modeladas como una secuencia de estados, considerando para ello los objetos empleados, y la posición del propio usuario. La información proporcionada por los sensores y los objetos se emplea para modelar

una serie de estados estándar para el usuario, mediante árboles de decisión. El comportamiento del usuario se infiere en cada momento mediante los estados más recientes, obtenidos a través de los modelos [117].

El proyecto considerado como el entorno inteligente más completo en términos de variedad de tecnologías para la monitorización es el *PlaceLab* del Instituto Tecnológico de Massachusetts (MIT, del inglés *Massachusetts Institute of Technology*) [116]. Este proyecto del MIT hace uso de un sistema basado en servicios inteligentes para llevar a cabo análisis cuantitativos y cualitativos de las relaciones entre los factores ambientales y el comportamiento de los usuarios. El sistema consta de tres componentes principales: una red de sensores binarios para recoger información acerca del uso de los objetos, una herramienta consciente del contexto que los usuarios emplean para etiquetar sus actividades y una colección de algoritmos de aprendizaje supervisado empleados para modelar el comportamiento de los sujetos. Respecto a los sensores, este sistema consta de videocámaras, sensores de humedad y temperatura e interruptores magnéticos. El entorno está habitado por voluntarios durante diferentes períodos de tiempo [115].

2.5.2. Conjuntos de Datos Públicos

Los conjuntos de datos públicamente disponibles ofrecen a la comunidad investigadora una valiosa herramienta de trabajo para crear plataformas de experimentación estandarizadas con las que poder evaluar el rendimiento de los algoritmos de minería de datos. Lamentablemente, debido principalmente a cuestiones de privacidad y costes, este tipo de datos libremente disponibles no son abundantes.

La mayoría de los datos de lo que dispone la comunidad están tomados en un entorno totalmente controlado, típicamente un laboratorio. Es muy común usar como entorno de prueba la configuración de una cocina domestica, ya que es un lugar donde se realizan bastantes actividades bien conocidas y definidas. Sin embargo, aunque el empleo de un laboratorio permite obtener información acerca de cómo realizan la misma actividad diferentes personas, tiene el inconveniente de que los sujetos no tienen plena libertad para realizar las actividades que deseen, restando realismo a los datos obtenidos.

Sin embargo, hay casos donde sí se cumple ese criterio y los sujetos del experimento disponen de plena libertad para llevar a cabo las actividades que quieran. El conjunto de datos generado por el *PlaceLab* del MIT es un ejemplo de ello. En este entorno, diversas videocámaras grababan en todo momento para posteriormente etiquetar los datos capturados [115]. Para generar un segundo conjunto de datos usando este entorno, dos investigadores permanecieron en el laboratorio durante más de dos meses, aunque desafortunadamente sólo se pudieran anotar las actividades de uno de los sujetos [151].

El MIT también ha desarrollado un equipo portátil de sensores que permite la instalación de entornos de monitorización inalámbricos en cualquier emplazamiento de un modo sencillo [114]. Empleando este sistema y un numeroso conjunto de sensores se generaron dos conjuntos de datos durante un experimento que duró dos semanas. Los usuarios anotaban la actividad que estaban realizando cada quince minutos mediante un terminal inalámbrico que portaban en todo momento. Lamentablemente, la precisión de dicho método de anotación resultó ser muy pobre, por lo que los investigadores se vieron obligados a emplear las lecturas de los sensores para determinar las actividades reales de los sujetos en cada instante de tiempo [230].

Recientemente, en la Universidad de Amsterdam (UvA, del neerlandés *Universiteit van Amsterdam*) han empleado una serie de nodos inalámbricos denominados *de enjambr* diseñados por la empresa RFM [204] para monitorizar personas mayores en el ámbito de su propio hogar. Dichos nodos son dispositivos inalámbricos que se comunican mediante tecnología WiFi, su topología es flexible y disponen de protocolos especialmente diseñados para obtener un bajo consumo de energía. Los conjuntos de datos obtenidos mediante esta red se generaron usando sensores infrarrojos de movimiento, sensores magnéticos, sensores de nivel y sensores de presión [131]. El etiquetado de las actividades lo llevaron a cabo usando unos auriculares inalámbricos junto a una aplicación de reconocimiento de voz.

No obstante, todavía es difícil encontrar conjuntos públicos que contengan datos realistas (tomados fuera de un entorno totalmente controlado), sean completos y que contengan un etiquetado de calidad acerca de las actividades.

2.6. Soluciones Comerciales

El éxito actual de los sistemas automáticos de supervisión se puede apreciar en la atención que están atrayendo durante los últimos años, y en el interés que diferentes empresas están mostrando en el área.

Recientemente, tanto centros de investigación como empresas privadas han presentando diversas soluciones para la integración de sistemas de monitorización en los hogares, con el objetivo de supervisar de un modo automático y ubicuo a personas dependientes o discapacitadas.

La empresa *Healthsense* ha desarrollado y comercializa el sistema de monitorización remota *eNeighbor* [99]. Este sistema está enfocado al cuidado de las personas mayores, e incluye detectores de caídas, seguimiento de AVDs, sensores de movimiento y botones inalámbricos de pánico. Su objetivo es la identificación temprana de los cambios en los hábitos de los usuarios que puedan indicar algún tipo de problema de salud. Para poder modelar el comportamiento de los usua-

rios, establece niveles de actividad, y asocia cambios bruscos en dichos niveles con alteraciones en las rutinas diarias.

Un sistema parecido es el desarrollado por la empresa *GrandCare Systems* [224]. Aunque este sistema no emplea técnicas para el modelado automático del comportamiento de sus usuarios, combina diversas tecnologías como pantallas táctiles, sensores de presión y movimiento, dispositivos de telemedicina y un portal web para fomentar la socialización de sus usuarios.

Sonamba es el nombre del dispositivo comercializado por la empresa *Pom-devices* para la monitorización del bienestar de personas que tengan alguna condición discapacitante [192]. El dispositivo es similar a una tableta y permite la comunicación de los usuarios con sus familiares o cuidadores. A excepción de un botón de pánico, los sensores del sistema están todos integrados en el propio dispositivo.

Recientemente, la empresa *Assisted Living Services, Inc.* ha empezado a comercializar el sistema *Care Plus* [7]. Este producto combina aplicaciones móviles con sensores de carácter médico para poder supervisar de forma remota el estado de salud de las personas mayores. Además genera alertas de forma automática cuando sus sensores detectan alguna variación importante en la salud de los usuarios.

Además de empresas relativamente pequeñas, grandes grupos corporativos también se han interesado por el mercado que suponen los sistemas de monitorización remota. Un claro ejemplo es el caso de la empresa *Philips*, que es el líder en la comercialización de servicios de alerta médica. Su gama de productos *Lifeline* ofrece diversos dispositivos para controlar el estado de salud de los pacientes e incluso gestionar sus medicamentos. El servicio también incluye alertas automáticas para situaciones de riesgo, como por ejemplo, caídas [175].

Capítulo 3

Captura y Preprocesado de Datos

En este capítulo se detallan tanto el sistema de monitorización como el método de anotación empleados para capturar y etiquetar nuestros conjuntos de datos.

En primer lugar se presenta la red de sensores inalámbricos que se ha diseñado, construido y desplegado para generar los conjuntos de datos. Se detallan la arquitectura y topología del sistema, así como los diferentes tipos de sensores que se han integrado en el entorno de monitorización.

Después, se describe el método de anotación que se ha seleccionado para que nuestros usuarios especifiquen qué actividades han llevado a cabo durante la captura de los datos. Esta información es muy importante para poder generar los conjuntos de datos etiquetados necesarios para el aprendizaje supervisado.

Por último, indicamos cual es el primer paso en el preprocesado de los datos y la notación que se ha seleccionado para esta tesis doctoral.

3.1. Red Inalámbrica de Sensores

De entre los diferentes tipos de tecnologías de monitorización existentes, en este trabajo se han elegido las redes de sensores inalámbricos (WSN, por sus siglas en inglés) como mecanismo para observar el comportamiento de una persona mayor en su propio hogar. Dichas redes se componen de una colección de dispositivos autónomos, distribuidos físicamente por el espacio que se desea vigilar, denominados nodos. En nuestro caso, para convertir el domicilio de la persona mayor en un entorno sensado, se han empleado un tipo de nodos inalámbricos denominados de enjambre. Dichos nodos son dispositivos que se comunican mediante algún tipo de tecnología inalámbrica, y a los que se les puede acoplar cualquier sensor analógico o digital. Estas redes de sensores se denominan de enjambre debido a que su topología es flexible y no requieren de un número fijo de nodos para funcionar. La gran ventaja de estos nodos es que aceptan prácticamente cualquier tipo de sensor

lógico, por lo que empleando sensores sencillos y baratos se puede componer una red de tamaño significativo que sea capaz de monitorizar numerosos elementos de un mismo domicilio.

Tras considerar diversas soluciones comerciales para la creación de redes de sensores inalámbricos personalizadas, nos decantamos por los kits de desarrollo DM1810-434-DK ofrecidos por la empresa RFM [204] (Figura 3.1(a)). Dichos kits incluyen una interfaz de programación de aplicaciones (API, por sus siglas en inglés) muy completa y bien documentada, y su *firmware* propietario incluye un protocolo de red energéticamente eficiente para el encaminamiento y la planificación de los mensajes. Dicha tecnología de comunicación de bajo consumo, junto al mecanismo de reposo que incluyen por defecto los nodos, permite un ahorro muy significado de energía y una mayor autonomía del sistema. Los nodos de estos kits tienen una velocidad de transmisión de datos de 4,8 kb/s, lo cual es más que suficiente para nuestros sensores binarios. Además de los nodos, cada red inalámbrica incluye una estación base (Figura 3.1(b)), que funciona como puerta de enlace conectándose a un ordenador mediante un puerto USB. Al sistema se le pueden añadir nuevos nodos muy fácilmente, mediante un proceso de sincronización muy sencillo que tan sólo implica pulsar un botón. La red de sensores inalámbricos RFM permite trabajar tanto con sensores digitales como analógicos, mandando un evento cada vez que la entrada digital cambia de estado o cada vez que se supera un cierto umbral de la entrada analógica.

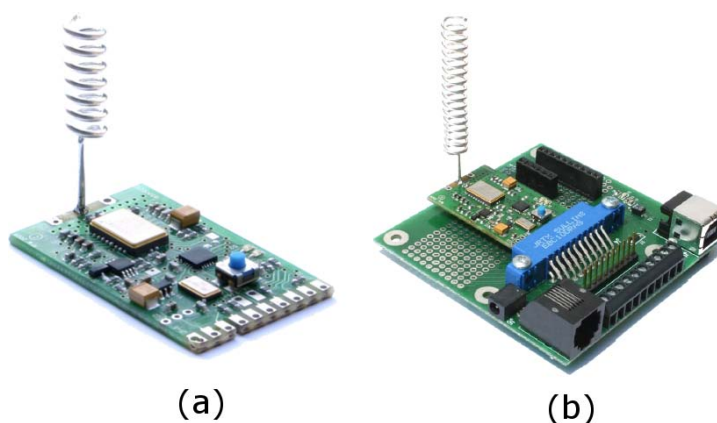


Figura 3.1: Nodo inalámbrico RFM DM 1810 (a) y nodo base RFM DM 1810 (b).

3.1.1. Tipos de Sensores Empleados

En nuestro sistema, los tipos de sensores elegidos para ser integrados en los nodos de la red fueron:

Sensores infrarrojos pasivos. Este tipo de sensores infrarrojos registran la radiación térmica humana y son capaces de detectar movimiento. Son especialmente precisos al registrar movimientos transversales al sensor (movimiento tangencial), es por ello que los usamos principalmente para detectar el paso del usuario a través de las puertas. También se emplean para detectar movimiento en áreas muy definidas, tales como la ducha o el lavabo.

Sensores magnéticos. Este tipo de sensores se componen de un interruptor eléctrico activado por un campo magnético. Los contactos del interruptor cambian de estado en presencia de un campo magnético, sirviendo como entrada digital de los nodos. Estos sensores se han usado para detectar la apertura tanto de puertas, como de armarios o frigoríficos aptos para su instalación.

Sensores de nivel. Este tipo de sensor es esencialmente un detector de nivel de líquidos, el cual indica si el nivel de un depósito ha rebasado o no un punto establecido. Funcionan de un modo mecánico, mediante un pequeño flotador, y se instalaron en las cisternas de los baños, para detectar el uso del inodoro.

Sensores de presión. Estos sensores tienen forma de esterilla y detectan de manera mecánica si se produce un punto de presión sobre ellos. Debido a su diseño, son capaces de detectar el peso de una persona dentro de un área relativamente amplia. Pueden registrar cuándo una persona se sienta o tumba encima, es por ello que resultan especialmente útiles para identificar el uso de la cama o los sofás.

Sensores de corriente eléctrica. Este tipo de sensores detectan la corriente eléctrica (continua o alterna) que circula por un cable, y generan una señal proporcional a ella. En nuestro sistema se emplearon sensores de efecto Hall que se activaban cuando la corriente sobrepasaba un cierto umbral. Estos sensores se emplearon para detectar el uso de ciertos electrodomésticos, como la tostadora o el microondas.

3.1.2. Arquitectura del Sistema

La arquitectura del sistema de monitorización fue diseñada de forma modular, tomando como premisa la escalabilidad y portabilidad de los diferentes componentes. Existen tres componentes principales en el diseño: la red de sensores, el servidor de administración centralizada y una plataforma móvil, la cual está compuesta por terminales telefónicos y cuya finalidad es recibir las alertas y mensajes que el sistema emita.

El mecanismo elegido para la comunicación entre los diferentes componentes es el patrón publicación/suscripción. Publicación/suscripción es un paradigma de comunicación asíncrona donde el emisor no indica específicamente el receptor de sus mensajes, sino que las entidades implicadas se pueden dividir en dos grandes roles: editores y suscriptores [79]. Los editores producen la información sin planear el envío de los mensajes de forma directa a suscriptores específicos, en

su lugar, publican dicha información en uno o más canales de datos que se encuentren disponibles. De forma similar los suscriptores consumen la información, recibiendo únicamente los mensajes que son de interés de los canales a los que están suscritos.

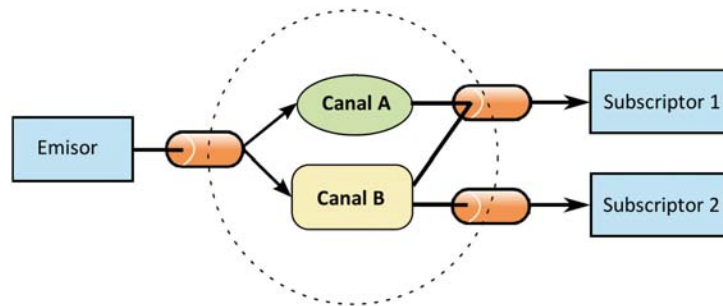


Figura 3.2: Esquema de comunicación publicación/suscripción.

Para que se dé este comportamiento, un sistema publicación/suscripción se compone de los siguientes elementos (ver Figura 3.2):

- **Productor de información.** Entidad emisora, que posee la información que hay que difundir. Dicho emisor publica la información a través de canales, sin tener en cuenta quién está interesado en recibirla.
- **Consumidor de información.** Entidad interesada en recibir información. El receptor se suscribe únicamente a los canales que diseminan la información que le interesa.
- **Mediador.** También conocido como *broker*, es una entidad de intermediación que está entre el productor y el consumidor. Recibe información de los emisores y peticiones de suscripción de los consumidores. También se encarga de encaminar la información publicada a los destinatarios suscritos al canal correspondiente.
- **Canal.** Es cada uno de los conectores entre los emisores y los receptores de información. El canal determina las propiedades de la información que se distribuye a través de él. Los canales modelan una relación de uno a muchos entre productores y consumidores.

Existen tres tipos principales de patrones publicación/suscripción que pueden ser usados para crear un mecanismo para enviar los datos: “Basado en el canal”, “Basado en la difusión”, y “Basado en el contenido” [77].

- **Basado en el canal.** Los elementos del sistema intercambian información a través de un conjunto de canales lógicos, normalmente conocidos de antemano. Todos los suscriptores de uno de estos canales recibirán todos los mensajes publicados en dicho canal.
- **Basado en la difusión.** Todo mensaje se transmite a todo el sistema. Cada una de las entidades que reciba la información deberá inspeccionarla y en el caso de que esté suscrita al tema del mensaje, procesarla.
- **Basado en el contenido.** En este enfoque, las suscripciones están directamente relacionadas con contenido específico de información, ofreciendo así el modelo publicación/suscripción más flexible y versátil.

La transmisión de mensajes unidireccionales que caracteriza los modelos publicación/suscripción proporciona una comunicación asíncrona entre los distintos elementos. Los emisores no se bloquean mientras producen eventos, y de forma similar los suscriptores pueden recibir datos asíncronamente mientras desempeñan otras tareas concurrentes.

De entre las diferentes implementaciones existentes del modelo publicación/suscripción, la que nosotros elegimos para integrar en nuestro sistema es el protocolo MQTT, acrónimo de “*MQ Telemetry Transport*”. MQTT es un protocolo abierto de mensajes publicación/suscripción [171] muy simple y ligero, diseñado para dispositivos limitados y de bajo ancho de banda, alta latencia o redes poco fiables. Los principios de diseño se basan en intentar minimizar el ancho de banda de la red y los recursos requeridos por el dispositivo, mientras garantiza la fiabilidad y cierto grado de garantía en la entrega. Una ventaja importante de MQTT es que la complejidad del sistema reside en la implementación del *broker*, permitiendo clientes simples y ligeros. Para asegurar una entrega adecuada de los mensajes, MQTT soporta diferentes niveles de calidades de servicio soporta básica calidad de servicio (QoS, por sus siglas en inglés) [42]. Las

calidades de servicio son las tecnologías que garantizan la transmisión de cierta cantidad de información en un tiempo dado. Dependiendo de la fiabilidad requerida por el sistema en la transmisión de datos, MQTT se puede configurar para emplear tres niveles de QoS. Sin embargo, un nivel de QoS más alto implica mayores requerimientos de latencia y ancho de banda, y para nuestro sistema una calidad QoS0 ha dado resultados muy positivos.

La entidad de intermediación que emplea MQTT también supervisa el estado de la conexión a través de un mecanismo que establece el intervalo máximo de tiempo que puede transcurrir entre dos mensajes recibidos por el cliente.

En la Figura 3.3 puede verse un esquema general de la arquitectura del sistema, en donde el protocolo MQTT define la comunicación entre los diferentes

El diagrama ilustra la arquitectura de la red de sensores inalámbricos (WSN) y su conexión con Internet. A la izquierda, un grupo de sensores (Nodo Sensor) se comunica con una pasarela (Pasarela) dentro de la WSN, operando a 433,92 Mhz RF COMM. La pasarela se conecta a un servidor (Servidor) que contiene una base de datos MySQL, una aplicación (App) y un broker (Broker). El servidor también se conecta a un proxy (Proxy) que actúa como intermediario entre el servidor y Internet. Finalmente, Internet conecta con los dispositivos móviles (Nodo MQTT Móvil) que se comunican con el proxy.

Figura 3.3: Diseño de la arquitectura del sistema.

Para poder evaluar el sistema propuesto en esta tesis doctoral, se han generado tres conjuntos de datos, para ello, se han instalado sistemas de monitorización en tres domicilios diferentes. El perfil genérico de usuario al que este trabajo se enfoca es el de una persona mayor, que viva sola en su hogar y que disponga de un familiar o cuidador que pueda realizar las labores de supervisión.

Para crear nuestros conjuntos de datos se ha contado con la colaboración de tres voluntarios. Estas personas fueron elegidas de acuerdo a su diferente nivel de autonomía, con el objetivo de introducir variabilidad en nuestros datos. Los perfiles de estas personas son:

1. Un hombre adulto autónomo e independiente (Conjunto de datos ‘OrdenezA’).
2. Una mujer mayor diagnosticada con la enfermedad de Parkinson (Conjunto de datos ‘OrdenezB’).

Tabla 3.1: Configuración de las viviendas monitorizadas.

| Conjunto de datos | OrdonezA | OrdonezB | OrdonezC |
|-------------------|----------|-------------|----------|
| Usuario | Hombre | Mujer | Mujer |
| Vivienda | Piso | Apartamento | Piso |
| Habitaciones | 4 | 2 | 5 |
| Duración | 14 días | 25 días | 21 días |
| Sensores | 12 | 12 | 12 |

3. Una mujer mayor independiente sin trastornos crónicos (Conjunto de datos ‘OrdonezC’).

Los diseños de planta de cada uno de los domicilios, en donde se indica la ubicación de los sensores puede apreciarse en las Figuras 3.4, 3.5 y 3.6.



Figura 3.4: Diseño de planta del domicilio empleado para generar el conjunto de datos ‘OrdonezA’.

Cada una de las viviendas tiene diferente tamaño y disposición, lo que junto a otra serie de características hace que cada entorno sea único. Es por esto que no es posible desplegar la misma red de sensores en todas las viviendas. En función del tipo y cantidad de electrodomésticos, el número de habitaciones y la estructura general de la vivienda, variaban los elementos a monitorizar y, por tanto, se escogió una configuración de sensores u otra. Un resumen de la distribución de sensores por vivienda se muestra en la Tabla 3.2.

3.2. Anotación de Actividades

Un aspecto importante dentro del aprendizaje supervisado es el correcto etiquetado de los datos de entrenamiento. En el caso del problema de reconocimien-



Figura 3.5: Diseño de planta del domicilio empleado para generar el conjunto de datos 'OrdenezB'.



Figura 3.6: Diseño de planta del domicilio empleado para generar el conjunto de datos 'OrdenezC'.

to de actividades esto se traduce en anotar adecuadamente qué actividad se ha realizado en cada instante de tiempo. El proposito de etiquetar las secuencias de

Tabla 3.2: Distribución de sensores por vivienda.

| Conjunto de datos | OrdonezA | OrdonezB | OrdonezC |
|---------------------------------|-------------------|-----------------------|-----------------------|
| Sensores infrarrojos pasivos | Ducha | Ducha | Ducha |
| | Lavabo | Lavabo | Lavabo |
| | Vitrocerámica | Puerta de la cocina | Puerta de la cocina |
| | | Puerta del salón | Puerta del salón |
| Sensores magnéticos | | Puerta del dormitorio | Puerta del dormitorio |
| | Puerta de entrada | Puerta de entrada | Puerta de entrada |
| | Armario baño | Frigorífico | Frigorífico |
| | Alacena | Armario medicinas | Alacena |
| Sensores de nivel | Frigorífico | | |
| | Inodoro | Inodoro | Inodoro |
| | Cama | Cama | Cama |
| | Silla | Sofá | Sofá |
| Sensores de corriente eléctrica | Tostadora | Microondas | Microondas |
| | Microondas | | |

datos de los sensores es poder identificar de forma correcta la actividad que se corresponde con cada secuencia de eventos generada por los sensores. Los datos correctamente etiquetados pueden entonces ser empleados para generar un modelo que defina una relación entre sensores y actividades, y sirva para clasificar nuevas secuencias de datos no observadas.

Sin embargo, en el ámbito del reconocimiento de actividades, tradicionalmente se ha prestado poca atención al problema de anotar las actividades del conjunto de datos. En la literatura, el método más común para etiquetar las actividades es solicitar al propio usuario que indique cuándo y durante cuánto tiempo lleva a cabo cada actividad [148, 230, 186]. En otros casos, las actividades son realizadas por el usuario de acuerdo a un orden o plan preestablecido, de modo que las actividades son identificadas incluso antes de la captura de los datos [49, 161]. Existen también ejemplos de trabajos donde el etiquetado de las actividades se ha realizado teniendo en cuenta únicamente las secuencias de eventos de los sensores, sin ningún tipo de información adicional que ayude en su identificación [256].

Ninguno de estos enfoques puede considerarse totalmente adecuado para todos los tipos de sistema que existen dentro del área del reconocimiento de actividades. Por ejemplo, cuando el sistema de monitorización está enfocado a supervisar personas con desórdenes cognitivos, no es razonable esperar que dichas personas sean capaces de anotar sus actividades diarias de forma adecuada. Y etiquetar los datos a mano, teniendo en cuenta que se deben revisar manualmente todos los datos generados por los sensores, puede ser demasiado costoso para según qué dominios.

En la literatura aparecen alternativas diferentes, que han sido propuestas y evaluadas en el área del reconocimiento de actividades. En un estudio, Szewczyk

et. al. detallaron los cuatro métodos más presentes para la anotación de actividades en entornos monitorizados [225]:

Datos secuenciales de los sensores. Este método se aplica en las situaciones donde únicamente se dispone de los datos secuenciales generados por la red de sensores, sin ningún etiquetado de los datos o información adicional. Es común hacer uso de un mapa o diagrama de la red de sensores y etiquetar los datos haciendo correlaciones entre cada actividad y un sensor o grupo de sensores. Es el mecanismo de anotación más costoso en términos de tiempo y además el más propenso a incluir errores humanos, y por tanto implica obtener un etiquetado de peor calidad. Sin embargo, es el método menos invasivo y el más respetuoso con la privacidad de los usuarios.

Datos secuenciales de los sensores + Diarios de actividad del usuario. Este método complementa los datos de los sensores con documentación generada por el propio usuario acerca de qué actividades ha llevado a cabo en cada instante de tiempo. Es muy común proveer al habitante de la casa con algún tipo de cuaderno o diario de anotaciones especialmente diseñado para escribir las actividades que realiza, cada media hora o cada hora, desde que se despierta hasta que se acuesta. Solicitar al usuario que rellene un documento todos los días implica un mayor nivel de invasividad, sin embargo, es percibido por los usuarios como más respetuoso que otros mecanismos existentes[230]. Sin embargo, los cuadernos en papel requieren poca dedicación por parte del usuario y son muy sencillos de emplear, siendo especialmente útiles cuando es necesario supervisar a personas mayores, las cuales habitualmente no están familiarizadas con tecnologías más actuales.

Visualización del entorno. Este método implica la utilización de alguna herramienta que permita la observación del ambiente monitorizado, ya sea de un modo directo mediante videocámaras o mediante algún tipo de tecnología de soporte, tales como los entornos virtuales tridimensionales. Algunos autores han hecho uso de simuladores virtuales para modelar el espacio y poder así obtener información acerca de los movimientos y actividades de los usuarios [225]. Las grabaciones se analizan a posteriori para llevar a cabo el etiquetado de los datos. Sin embargo, la invasividad de este tipo de tecnologías es bastante elevada, y tiene un alto porcentaje de rechazo por parte de los usuarios.

Visualización del entorno + Entrevistas con el usuario. Este último método supone implicar al usuario en la visualización e interpretación de la información visual obtenida del entorno. Es el mecanismo mediante el cual se obtiene los conjuntos de datos mejor etiquetados y más precisos, pero del mismo modo es el método más invasivo y, por lo tanto, es difícilmente aplicable en la obtención de datos reales.

Los datos obtenidos para esta tesis doctoral se han etiquetado empleando diarios de actividad y revisando los datos en bruto generados por los sensores. Los usuarios del sistema rellenaban diariamente una sencilla plantilla anotando qué

actividades habían realizado y en qué momento de día. Tomando como base dicho cuaderno de anotaciones, los investigadores etiquetaban manualmente los conjuntos de datos. Se optó por dicho método debido a que combina una baja invasividad y una buena calidad de etiquetado [225], por lo que es considerado el método de anotación más indicado para nuestro dominio. Además, debido a que el usuario objetivo del sistema es una persona mayor, emplear diarios de actividad en papel no supone la barrera tecnológica que, a priori, podrían significar otro tipo de tecnologías. Para poder evaluar cómo es percibido este sistema por sus usuarios se llevó a cabo un estudio mediante cuestionarios y entrevistas, y cuyos resultados se incluyen en el Apéndice C.

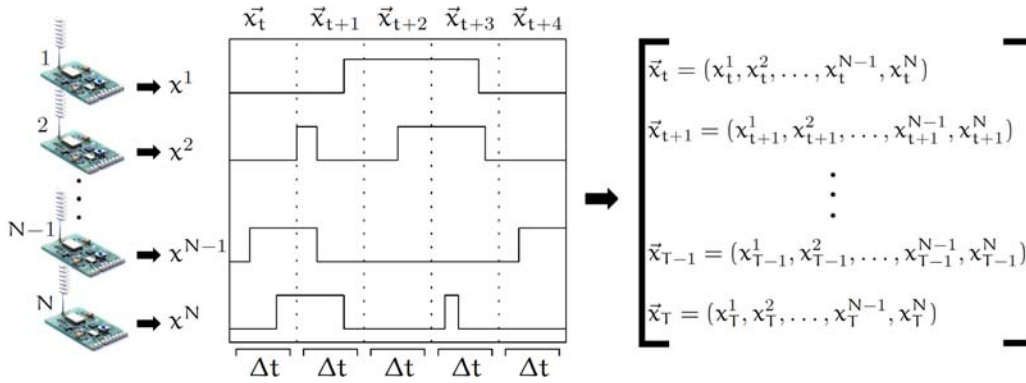


Figura 3.7: Segmentación de los datos.

La lista de posibles actividades incluidas como etiquetas para los conjuntos de datos se definió tomando como base los trabajos de Munguia *et. al.* [230] y de Kasteren *et. al.* [127]. Para nuestro dominio, se seleccionaron las AVD que incluyesen información relevante sobre la movilidad y cuidado diario de las personas, y que además fuesen adecuadas a la edad y perfil de nuestros usuarios objetivo. Las actividades que finalmente se incluyeron para etiquetar los datos son: ‘Aseo’, ‘Ducha’, ‘Usar el inodoro’, ‘Desayunar’, ‘Almorzar’, ‘Cenar’, ‘Tomar un refrigerio’, ‘Salir’, ‘Dormir’, ‘Ocio’. Dentro de la categoría ‘Ocio’ se enmarcan actividades tales como ver la televisión, leer o hacer uso de un ordenador.

3.3. Notación y Discretización

Para proporcionar un formato temporal adecuado al conjunto de datos, se empleó un mecanismo de ventana deslizante sobre los datos obtenidos a través de los sensores. El flujo de datos generado por dichos sensores binarios se discretizó a intervalos regulares, con una granularidad de tiempo Δt . Cada evento de los sensores se expresa como x_t^i , indicando si el sensor i se ha activado al menos

una vez durante el intervalo definido entre el instante t y el instante $t + \Delta t$, donde $x_t^i \in \{0, 1\}$. Por tanto, en un entorno monitorizado donde existan N sensores binarios instalados, se define un vector $\vec{x}_t = (x_t^1, x_t^2, \dots, x_t^{N-1}, x_t^N)^T$, compuesto por valores lógicos, en cada uno de los intervalos de tiempo definidos.

La Figura 3.7 resume la representación de datos empleada, en donde cada intervalo de tiempo se corresponde a una única instancia de datos. La actividad que ha realizado el usuario en cada instante de tiempo t , y que define la clase de la instancia, se expresa mediante $y_t \in \{1, \dots, Q\}$, para Q posibles actividades. Es por ello, que la tarea de clasificación supervisada a la cual nos enfrentamos es la de realizar una asignación entre una secuencia de observaciones $\mathbf{x} = \{\vec{x}_{t1}, \vec{x}_{t2}, \dots, \vec{x}_T\}$ y una secuencia de actividades $\mathbf{y} = \{y_{t1}, y_{t2}, \dots, y_T\}$ para un total de T intervalos de tiempo.

Mediante este proceso se obtienen las instancias de datos con las que se realiza el aprendizaje de los métodos presentados en esta tesis doctoral, los cuales se detallan en los siguientes capítulos.

Capítulo 4

Reconocimiento de Actividades

En el presente capítulo detallamos nuestra propuesta para el reconocimiento de actividades en entornos monitorizados mediante el uso de modelos híbridos, los cuales combinan una parte generativa que controla la secuencialidad de las actividades, con una parte discriminativa que interpreta los datos de los sensores. Además, presentamos un mecanismo de aprendizaje basado en inferencia bayesiana orientado a la aplicación de los HMM en nuestro dominio, con el fin de solventar varias de las deficiencias planteadas por los métodos de inferencia clásicos.

El reconocimiento de las actividades humanas posee ciertas características que lo convierten en un problema no trivial: cada usuario realiza cada actividad específica de un modo personal y diferente, la duración de las actividades es habitualmente desconocida y los datos generados por los sensores son ruidosos y tienen un comportamiento no determinista. Sin embargo, existe un tipo de aprendizaje basado en el modelo de datos, conocido como *data-driven*, que ha demostrado ofrecer un buen comportamiento y ser capaces de manejar bien la incertidumbre causada por las características de este dominio [46].

Dentro del concepto *data-driven* se engloban aquellas técnicas estadísticas y probabilísticas que son capaces de inferir modelos de actividad, empleando conjuntos de datos representativos de dichas actividades, a través de técnicas de minería de datos y aprendizaje automático. Estas técnicas son capaces de tomar una decisión informada acerca de qué actividad está realizando el usuario basándose en la evidencia que los sensores proporcionan. Según su enfoque, estos métodos de aprendizaje pueden clasificarse en dos grupos: aprendizaje generativo y aprendizaje discriminativo [182]. En el enfoque generativo, la finalidad es obtener un esquema que contenga la descripción del espacio de características del problema, normalmente mediante un modelo probabilístico. En el enfoque discriminativo, el objetivo es generar un modelo que establezca una correspondencia entre los atributos del problema (datos de los sensores) y una clase (actividad a recono-

cer). Dentro de los métodos discriminativos se encuentran populares algoritmos de aprendizaje supervisado, tales como las redes de neuronas artificiales (RNA), las máquinas de soporte vectorial (*Support Vector Machines* o SVM en inglés) o los campos aleatorios condicionales (*Conditional Random Field* o CRF en inglés).

De entre los diferentes modelos generativos, los modelos ocultos de Markov son una técnica muy popular, ampliamente presente en la literatura, que han sido aplicados con éxito en muchos problemas de modelado de datos secuenciales, tales como el reconocimiento de voz [203], reconocimiento de escritura [10] y análisis de secuencias biológicas [254]. En el dominio del reconocimiento de actividades, es posible utilizar estas técnicas para modelar el comportamiento humano como una secuencia probabilística de transiciones, en donde cada estado representa una actividad.

Un importante artículo de referencia relativo a los HMM es el presentado por Rabiner en 1989 [197]. En este trabajo se detalla la topología de estos modelos y los tres problemas fundamentales que los caracterizan (evaluación, descifrado y aprendizaje). Una de las primeras propuestas para reconocer acciones humanas mediante HMM fue la presentada por Yamato *et. al.* [261], en donde el modelo se generó a partir de secuencias de imágenes.

A pesar de su popularidad, los HMM también tienen sus limitaciones. Estos modelos no son capaces de modelar las dependencias transitivas que se den en los datos de aprendizaje, debido a que aceptan una hipótesis de independencia muy estricta entre los datos que componen las observaciones del modelo. Además, los algoritmos de máxima expectación que típicamente se han empleado para aprender el modelo, tienen un rendimiento muy condicionado a los valores iniciales de los parámetros, por lo que dan lugar a una exploración no uniforme del espacio de soluciones [96].

Una limitación de los modelos generativos es el hecho de necesitar una cantidad significativa de datos para poder modelar de forma completa las distribuciones probabilísticas que requieren. Así, los modelos discriminativos suponen una alternativa a ese problema, puesto que no buscan representar el problema probabilísticamente, sino que buscan una solución que resuelva directamente la tarea de clasificación. Su objetivo es encontrar un límite o hipersuperficie que permita dividir las diferentes partes del espacio de características en grupos o clases.

Muchas técnicas discriminativas toman en cuenta explícitamente el hecho de que para poder llevar a cabo la tarea de clasificación, los puntos de espacio que realmente resultan más interesantes son aquellos que determinan la hipersuperficie que divide los datos en diferentes clases. Aquellos datos que se encuentran más alejados de dichos límites, y cuya clasificación es consecuentemente más sencilla, son los menos influyentes en el rendimiento de los modelos. Los datos más próximos a los hiperplanos que delimitan las clases, y que por tanto son los más interesantes en la tarea de clasificación, son denominados *vectores de soporte* y

son la base de las técnicas de aprendizaje supervisado conocidas como máquinas de soporte vectorial. Las SVM han sido también empleadas en el dominio del reconocimiento de actividades. Dentro del proyecto CareMedia se ha usado estos algoritmos discriminativos para localizar y reconocer interacciones sociales en un centro de atención, empleando múltiples sensores, audio y vídeo [44]. Ravi *et. al.* demostraron cómo las SVM funcionan de forma bastante precisa tanto individualmente como integradas dentro de conjuntos de clasificadores [201]. Empleando acelerómetros, extrajeron conjuntos de características de los movimientos de los usuarios, y usando conjuntos de clasificadores con diferentes topologías etiquetaron dichas actividades. Los conjuntos de clasificadores que incluían SVM, obtuvieron los mejores resultados en las configuraciones experimentales más exigentes.

Nuestro problema de reconocimiento de actividades de la vida diaria se enfoca en determinar la secuencia más probable de actividades que haya generado el flujo de datos recibido de los sensores. El problema de seleccionar la secuencia de hipótesis más probable puede formalizarse como un HMM. De hecho, Kasteren *et. al.* demostraron que tanto el modelo oculto de Markov, como diversas variantes de dicho modelo, pueden aplicarse con éxito al problema [132].

Sin embargo, los modelos discriminativos también han ofrecido buenos resultados en problemas de aprendizaje de patrones similares al nuestro [154]. Concretamente, modelos bien conocidos como las redes de neuronas artificiales, proporcionan un aprendizaje discriminativo que presenta varias ventajas frente a los algoritmos generativos:

- Pueden modelar vectores de características de gran dimensionalidad.
- Consideran los datos de entrada en su conjunto, pudiendo ser empleados para hacer un modelado no paramétrico de sus distribuciones de probabilidad [21].
- No necesitan realizar ninguna asunción de independencia entre atributos.
- La salida de estos métodos son valores que representan probabilidades, simplificando la tarea de combinar fuentes de datos diferentes mediante técnicas de integración [137].

En este capítulo se detalla cómo abordar los problemas que plantean los modelos generativos en la tarea de reconocimiento de AVD, cuando se trabaja con un espacio de características como el que nuestra red de sensores nos ofrece.

Para abordar dicho problema, en esta tesis doctoral se proponen dos enfoques: (1) el empleo del aprendizaje discriminativo sobre nuestro espacio de características haciendo uso de clasificadores bien conocidos y modelos híbridos, y (2) el uso

de otro tipo de algoritmos para llevar a cabo el proceso de inferencia sobre los modelos generativos que ya han demostrado funcionar bien sobre nuestro dominio, los HMM.

En el siguiente apartado se hace hincapié en las diferencias que existen entre los dos tipos de modelos que hemos aplicado a nuestra tarea de reconocer AVDs. A continuación, se incluye una descripción general acerca de los modelos ocultos de Markov, puesto que son base y parte fundamental de nuestro enfoque. Después se detalla nuestra propuesta discriminativa para este problema, en donde planteamos, aplicamos y comparamos diferentes modelos híbridos con nuestros conjuntos de datos. A continuación se detalla nuestra alternativa al clásico algoritmo de esperanza-maximización para llevar a cabo el aprendizaje de los HMM en nuestro dominio, específicamente mediante técnicas bayesianas. Por último incluimos una comparativa entre los diferentes métodos aquí propuestos, junto a las conclusiones de esta parte de la tesis doctoral.

4.1. Modelos Generativos vs Modelos Discriminativos

Formalmente, la diferencia existente entre los métodos generativos y los métodos discriminativos reside en el modo en que cada uno de ellos aprenden los parámetros del modelo. Ambos enfoques buscan maximizar la probabilidad condicional $p(\mathbf{y}_{1:T}|\mathbf{x}_{1:T})$ para cada secuencia de observaciones, con el objetivo de determinar la secuencia óptima de estados que le corresponda. Pero cada modelo emplea unos criterios diferentes para la estimación de los parámetros.

En el caso de los métodos generativos, la estimación de los parámetros está enfocada a maximizar la distribución de probabilidad conjunta $p(\mathbf{y}_{1:T}, \mathbf{x}_{1:T})$. El muestreo desde este tipo de distribución permite la creación de nuevos ejemplos en el espacio de características de los datos, y de ahí viene el nombre de modelos generativos.

Por el contrario, en los modelos discriminativos la estimación de los parámetros se enfoca a maximizar la distribución de probabilidad condicional $p(\mathbf{y}_{1:T}|\mathbf{x}_{1:T})$, intentando así, resolver directamente el problema de determinar la probabilidad a posteriori de las clases. Así, se centran en identificar los límites del espacio de características que delimitan las clases, en vez de intentar modelar las propiedades individuales de cada clase.

En el ámbito del reconocimiento de actividades, los modelos discriminativos se consideran interesantes por varias razones. Por ejemplo, estos métodos permiten técnicas de aprendizaje incremental y no supervisado [111], incorporan mecanismos para tratar con datos no completos y son útiles para la construcción mo-

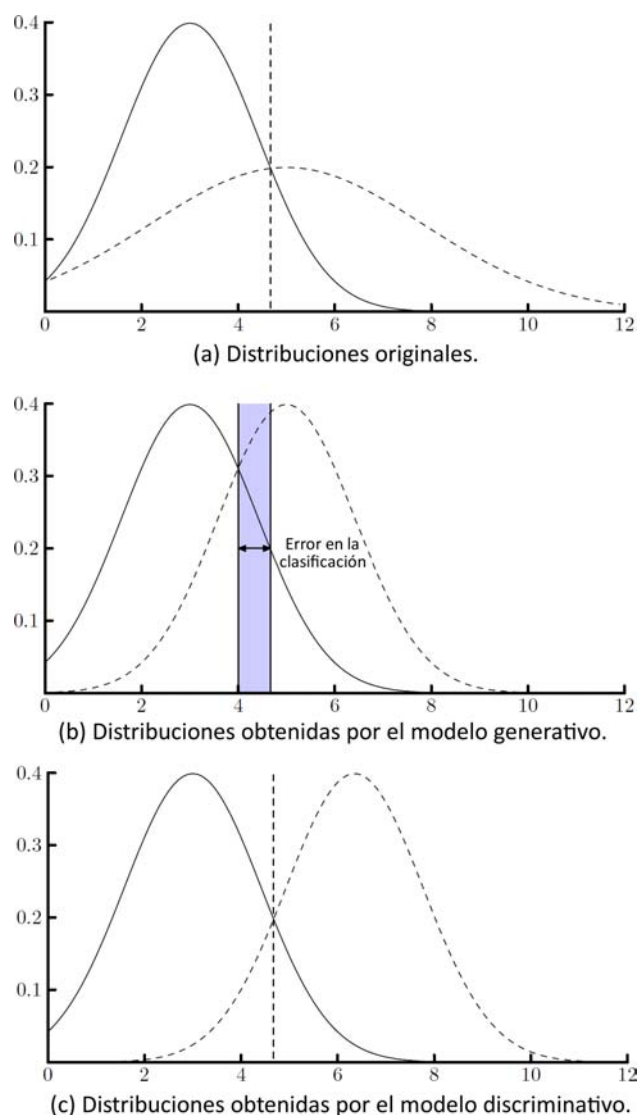


Figura 4.1: Ejemplos de asunciones incorrectas sobre el modelo propuesto por Kasteren *et. al.* [129].

dular de soluciones complejas, tales como los clasificadores jerárquicos. Además, permiten incorporar fácilmente probabilidad a priori en el proceso de aprendizaje.

Los métodos discriminativos, por su parte, permiten establecer límites de forma flexible en el espacio de características de los datos, ofreciendo un rendimiento en la tarea de clasificación a menudo superior que los obtenidos mediante técnicas puramente probabilísticas o generativas[118, 123]. En el dominio del reconocimiento de actividades, estas técnicas han demostrado ser capaces de llevar a cabo un aprendizaje explícito de los atributos discriminativos de una actividad en

particular [235].

La diferencia entre el aprendizaje generativo y el discriminativo se hace más evidente cuando basamos los modelos en suposiciones incorrectas [123]. Los clasificadores a menudo incluyen suposiciones incorrectas acerca del modelado de los datos, ya sea porque desconocemos la auténtica distribución de los datos, o porque la elección de la distribución da lugar a menos valores de los parámetros, requiriendo por tanto menos datos para aprender dichos parámetros de forma precisa.

Kasteren *et. al.* propusieron un ejemplo para ilustrar el impacto que puede tener basar los modelos en suposiciones incorrectas [129]. Se consideran dos clases A y B , en donde sus variables aleatorias x se encuentran independiente e idénticamente distribuidas, y donde ambas clases tienen la misma probabilidad de ocurrir ($p(y = A) = p(y = B) = 0,5$). Cada punto nuevo se asigna a una u otra clase basándose en el cálculo de la probabilidad a posteriori que ofrece la Regla de Bayes:

$$p(x|y) = \frac{p(x|y)p(y)}{p(x)} \quad (4.1)$$

Las variables aleatorias de la clase A ($p(x|y = A)$) están modeladas de acuerdo a una distribución gaussiana de parámetros $\mu = 3$ y $\sigma = 1$. Los datos de la clase B siguen también una distribución gaussiana, pero este caso con parámetros $\mu = 5$ y $\sigma = 2$ (ver Figura 4.1(a)).

En dicho estudio se realiza una comparación, empleando aprendizaje generativo y discriminativo, para modelar la distribución de las observaciones $p(x|y)$, a través de una distribución gaussiana con una desviación estándar fija $\sigma = 1$. Ya que la desviación estándar es fija, los parámetros a aprender son únicamente las medias de las distribuciones gaussianas.

En el caso del modelo generativo, los autores aprenden las medias maximizando la probabilidad conjunta $p(x, y)$. Esto es equivalente a maximizar la distribución de las observaciones $p(x|y)$ de manera independiente. Los parámetros son ajustados mediante estimación por máxima verosimilitud calculando la media de los puntos de entrenamiento disponibles. La Figura 4.1(b) muestra la distribución gaussiana resultante, tras el proceso de aprendizaje. Se puede apreciar cómo con el aprendizaje generativo se infieren correctamente las medias de las distribuciones, pero debido a suposiciones incorrectas sobre el modelo (se asume que la desviación estándar es igual a 1) existe un error en la clasificación.

En el caso del modelo discriminativo, los parámetros son estimados maximizando la probabilidad *a posteriori* $p(x|y)$ directamente. No existe una solución en forma cerrada para maximizar dicha función, por lo que algún método numérico como el descenso del gradiente debe emplearse. La Figura 4.1(c) muestra la distribución gaussiana que se obtiene para este caso. Se puede apreciar cómo el

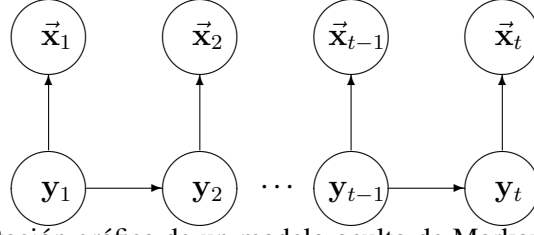


Figura 4.2: Representación gráfica de un modelo oculto de Markov. Los nodos y representan los estados ocultos y los nodos \vec{x} las observaciones.

modelo infiere una media incorrecta para la clase B ($\mu = 6, 35$), sin embargo el empleo de dicha media incorrecta da como resultado una clasificación óptima.

Este ejemplo que propusieron Kasteren *et. al.* demuestra cómo los modelos discriminativos son más robustos a la hora de tratar las violaciones a las suposiciones del modelo. Además, ilustra cómo los modelos generativos y los modelos discriminativos generan clasificadores que exploran el espacio de soluciones de forma diferente, por lo que una combinación de ambas técnicas puede ser potencialmente muy útil para nuestro dominio.

4.2. Modelos Ocultos de Markov

La meta al aplicar los HMM a nuestro dominio es reconocer qué AVD se llevan a cabo dada una secuencia de vectores de eventos generados por nuestro sistema de monitorización. Esto es, buscamos la secuencia de actividades $y_{1:T}$ que mejor justifique la secuencia de observaciones $\vec{x}_{1:T}$. En términos probabilísticos esto se corresponde con encontrar la secuencia $y_{1:T}$ que maximice la probabilidad $p(y_{1:T}|\vec{x}_{1:T})$.

Un HMM puede describirse como un modelo probabilístico generativo que modela un proceso definido en términos de un observación x_t y una variable oculta y_t para un sistema de tiempo discreto. En nuestro problema la observación viene determinada por el vector de eventos que generan los sensores en cada intervalo Δt , y la variable oculta es la AVD a reconocer.

Los modelos generativos ofrecen una representación explícita de dependencias empleando el valor de probabilidad conjunta sobre las variables ocultas y las observaciones. La función que expresa la probabilidad conjunta de las actividades y los datos de los sensores es la función $p(y_{1:T}, \vec{x}_{1:T}|\theta)$. Ésta es una función sobre el modelo de los parámetros θ , el cual es optimizado durante el proceso de aprendizaje.

La arquitectura de este modelo se define mediante dos suposiciones, representadas mediante el grafo dirigido que muestra la Figura 4.2.

- El valor de la variable oculta y_t (en el instante t), sólo depende de la variable oculta en el instante anterior y_{t-1} (*propiedad de Markov de primer orden* [197]).
- El valor de la variable observable \vec{x}_t (en el instante t), depende únicamente de la variable oculta y_t de ese mismo instante de tiempo.

Por tanto, la probabilidad conjunta está definida mediante la función

$$p(\mathbf{y}_{1:T}, \mathbf{x}_{1:T}) = p(\mathbf{y}_1) p(\mathbf{x}_1|\mathbf{y}_1) \prod_{t=2}^T p(\mathbf{y}_t|\mathbf{y}_{t-1}) p(\vec{\mathbf{x}}_t|\mathbf{y}_t). \quad (4.2)$$

En donde cada uno de los factores representa una distribución de probabilidad:

- Probabilidad incondicional de estado $p(\mathbf{y}_1)$. Es la probabilidad a priori, la probabilidad de empezar una secuencia de estados con el estado \mathbf{y}_1 . Está definida mediante una distribución multinomial, en donde los valores de sus parámetros se almacenan en el vector de estado inicial π y se representan $p(\mathbf{y}_1 = i) \equiv \pi_i$.
- Probabilidad condicional de transición $p(\mathbf{y}_t|\mathbf{y}_{t-1})$. Representa la probabilidad de pasar de un estado al siguiente.
- Probabilidad condicional de las observaciones o de emisión $p(\vec{\mathbf{x}}_t|\mathbf{y}_t)$. Las probabilidades de emisión caracterizan la verosimilitud de que el estado \mathbf{y}_t genere una observación $\vec{\mathbf{x}}_t$.

Para poder modelar de forma precisa y completa la distribución de las observaciones en nuestro dominio, se requeriría modelar todas las posibles combinaciones de los valores que definen las activaciones de los sensores. Este enfoque necesitaría 2^N atributos, donde N está definido por el número de sensores empleados. Incluso cuando el valor de N sea pequeño, el número de observaciones puede llegar a ser demasiado grande para ser un problema tratable. Tal y como demostraron Kasteren *et. al.* [132], la solución más plausible a este problema es el uso de la asunción "naïve" de independencia condicional, lo que implica que cada atributo se modela independientemente del resto. Usando esta asunción se requieren entonces únicamente N atributos para cada actividad. Sin embargo, representa una suposición muy estricta que muy probablemente no representa la distribución real de los datos, ya que es posible que dos sensores dependan el uno del otro para alguna actividad en particular. Pero emplear dicha asunción ha dado buenos resultados en otros trabajos [207] y se ha demostrado útil para nuestro dominio.

De acuerdo al tipo de sensores que empleamos, cada parámetro de las observaciones se modela como una distribución de Bernoulli independiente, en donde

μ_{in} representa el parámetro que modela el sensor n para la actividad i . Por tanto, la probabilidad condicional de las observaciones viene determinada por

$$p(\vec{\mathbf{x}}_t | \mathbf{y}_t) = \prod_{n=1}^N p(\mathbf{x}_t^n | \mathbf{y}_t) \quad (4.3)$$

en donde

$$p(\mathbf{x}_t^n = v | \mathbf{y}_t = i) = (\mu_{in})^v (1 - \mu_{in})^{1-v}. \quad (4.4)$$

Los valores de los parámetros de las observaciones se almacenan en la matriz \mathbf{B} . Por su parte, la probabilidad condicional de transición $p(\mathbf{y}_t = j | \mathbf{y}_{t-1} = i) \equiv a_{ij}$ está definida mediante una distribución multinomial cuyos valores se almacenan en la matriz \mathbf{A} .

Por tanto, un HMM es un modelo probabilístico temporal en donde en cada instante discreto de tiempo el proceso asume estar en un estado, y al recibir las observaciones el modelo cambia de estado según una distribución de probabilidad. Las funciones de densidad de probabilidad que lo caracterizan son tres: distribución sobre los estados iniciales parametrizada por $\pi = \{\pi_i\}$; la distribución de transición dada por $\mathbf{A} = \{a_{ij}\}$; y la distribución de las observaciones parametrizada por $\mathbf{B} = \{\mu_{in}\}$.

4.2.1. Inferencia

Para un modelo que contiene variables ocultas, como es el caso del HMM, la tarea de inferencia supone determinar qué secuencia de variables ocultas es la generadora de la secuencia de observaciones. Esto es, implica la selección de la secuencia de estados o camino que proporcione la máxima verosimilitud con respecto al modelo dado, o lo que es lo mismo, que maximice $p(\mathbf{y}_{1:T}, \mathbf{x}_{1:T})$. Aunque el número de posibles secuencias crece exponencialmente con la longitud de las mismas, dicho camino puede obtenerse de manera recursiva mediante el algoritmo de *Viterbi* [197]. Dicho algoritmo emplea programación dinámica para descartar un número de posibles secuencias en cada paso, resultando en una complejidad computacional $O(TQ^2)$ para cada secuencia, donde T define la longitud de la misma y Q el número de estados [197]. El algoritmo *Viterbi* se explica más detalladamente en el Apéndice A.

4.2.2. Aprendizaje de los Parámetros

El problema del aprendizaje en un HMM formalmente consiste en encontrar los valores para los parámetros $\theta = \{\pi, \mathbf{A}, \mathbf{B}\}$ que, dados unos datos de

entrenamiento $\mathbf{y}_{1:T}, \mathbf{x}_{1:T}$, maximicen la probabilidad $p(\mathbf{y}_{1:T}, \mathbf{x}_{1:T}|\boldsymbol{\theta})$. El algoritmo clásico para llevar a cabo este aprendizaje es el algoritmo *Baum-Welch* o *forward-backward* [197], que es un caso particular del algoritmo de esperanza-maximización (*Expectation-Maximization* o algoritmo EM). Este algoritmo permite entrenar tanto las probabilidades de transición \mathbf{A} como las probabilidades de emisión \mathbf{B} del modelo.

El algoritmo está definido mediante un proceso iterativo donde el conjunto de parámetros actuales $\boldsymbol{\theta}^t$ se emplea para computar la esperanza de la verosimilitud, y el nuevo conjunto de parámetros $\boldsymbol{\theta}^{t+1}$ es definido calculando estimadores de máxima verosimilitud de los parámetros mediante la realización de una maximización $\boldsymbol{\theta}^{t+1} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^t)$. Dicho proceso iterativo comienza con un conjunto de parámetros definidos de forma aleatoria, y los pasos de esperanza y maximización se alternan asegurando siempre que la probabilidad de los datos, dados los parámetros del modelo, no disminuye hasta alcanzar la convergencia a un valor $\boldsymbol{\theta}$ que asegure maximizar (localmente) $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^t)$.

Sin embargo, al ser un método de búsqueda en escalada, es un algoritmo numéricamente sensible y depende mucho de la estimación inicial de los parámetros del modelo. Al basarse en comparaciones locales, no explora exhaustivamente el espacio de búsqueda, y una estimación arbitraria de los parámetros iniciales muy posiblemente conducirá a soluciones subóptimas [96].

El algoritmo *Baum-Welch* se explica más detalladamente en el Apéndice A.

4.3. Modelos Híbridos

Tal y como se ha indicado anteriormente, el problema de reconocer AVD puede enfocarse mediante el uso de técnicas generativas que modelen las distribuciones subyacentes de las clases o bien mediante técnicas discriminativas que se centren en aprender los límites del espacio que definen dichas clases [210]. Ambos tipos de técnicas han sido empleadas en el marco del reconocimiento de actividades [46]. Sin embargo, de cara a afrontar los retos estadísticos que supone reconocer actividades humanas, donde diferentes tipos de acciones deben ser claramente diferenciadas y donde los algoritmos necesitan modelar las particularidades temporales de las actividades, estos dos tipos de métodos tienen deficiencias cuando se aplican independientemente.

En nuestro dominio, el espacio de características de nuestros sensores puede modelarse mediante el uso de modelos de Markov, un método generativo que ya ha ofrecido buenos resultados, pero que implica emplear suposiciones de independencia muy estrictas entre los atributos, y que no son realmente ciertas.

Por su parte, los algoritmos discriminativos más empleados, como las redes de neuronas o las máquinas de vectores de soporte, necesitan de algún método que

convierta nuestro problema de aprendizaje supervisado secuencial en un problema clásico de aprendizaje supervisado. Para llevar a cabo dicha transformación, en nuestro caso, hemos empleado un mecanismo de ventana deslizante, mediante el cual el flujo de datos de los sensores es segmentado en ejemplos de entrenamiento. El método de ventana deslizante construye un *clasificador de ventana* h_w que mapea una secuencia de datos de entrada, de longitud w , a un valor de salida y . Esto es, el clasificador h_w predice la actividad y_t usando la ventana $\langle \vec{x}_{t-w}, \vec{x}_{t-w+1}, \dots, \vec{x}_t \rangle$. El clasificador h_w se entrena convirtiendo cada ejemplo de entrenamiento (\vec{x}_i, y_i) en ventanas y aplicando el mecanismo de aprendizaje supervisado estándar. La secuencia de datos obtenida a través de los sensores se clasifica segmentando primero el flujo en ventanas, después aplicando el clasificador elegido h_w para predecir cada actividad y_t , y por último concatenando todas las salidas del clasificador para producir la secuencia de actividades predicha.

Usando el mecanismo de ventana deslizante es posible aplicar una multitud de clasificadores a nuestra secuencia de datos, pero un problema de este método es que no puede modelar las correlaciones entre actividades cercanas. Concretamente, las únicas correlaciones entre clases y_t que se pueden capturar haciendo uso de estos métodos son aquellas que se derivan directamente de los valores \vec{x}_t cercanos en la secuencia de datos. Si existe algún tipo de correlación entre dos actividades y_t que sea independiente de los valores \vec{x}_t , entonces este método no es capaz de identificarla.

En la presente tesis doctoral, hemos abordado estas limitaciones mediante la aplicación de un modelo híbrido, que combina las capacidades discriminativas de dos clasificadores bien conocidos, como son las redes de neuronas y las máquinas de vectores de soporte, junto a la capacidad de los HMM de capturar las relaciones temporales de los datos.

El modelo resultante se denomina modelo HMM híbrido. Dicho HMM híbrido modela las características temporales de los datos mediante la probabilidad condicional de transición del HMM y un clasificador es empleado para estimar la probabilidad de que una observación haya sido generada por cada uno de los posibles estados. Una ventaja importante de emplear estos modelos es que los algoritmos de inferencia y aprendizaje existentes para los HMM clásicos también pueden ser utilizados para los HMM híbridos sin apenas modificaciones, ya que un modelo híbrido se comporta esencialmente como un HMM estándar.

Distintos tipos de HMM híbridos han sido aplicados a diferentes problemas, pero sin duda, una de las propuestas más presentes en la literatura es la combinación de redes de neuronas con HMM [236]. Es posible integrar RNA con HMM debido a que la salida obtenida de estos clasificadores puede usarse para obtener la probabilidad a posteriori de las clases [26].

La mayor parte de la literatura sobre los modelos híbridos HMM/RNA se centra en el dominio del reconocimiento del habla. Desde los años 90, han sido varios

los autores que han empleado enfoques híbridos como una manera de superar los inconvenientes que plantean los HMM [26, 17, 203, 206, 80, 27]. Híbridos de las RNA dentro de los HMM, en el dominio del reconocimiento del habla, tiene por objeto proporcionar un entrenamiento más discriminativo, mejorar el modelado de la duración de los fonemas, y lo más importante, hacer un mayor uso de la información contextual. Los dos componentes del sistema híbrido pueden ser entrenados de forma independiente, pero varios autores han propuesto métodos para llevar a cabo una optimización combinada [16, 106, 237], la cual produce normalmente mejores resultados. El tipo de red neuronal empleado por la mayoría de estos sistemas híbridos es un perceptrón multicapa (*Multilayer Perceptron* o MLP en inglés) [169, 55]. Sin embargo, también ha existido un notable interés por parte de diversos autores en el uso de redes de neuronas recurrentes (*Recurrent Neural Network* o RNN en inglés) [215, 108, 221], debido a que el propósito principal de las redes neuronales en el problema del reconocimiento del habla es el de introducir información contextual al aprendizaje. Otros autores han enfocado su trabajo a combinar las redes de neuronas no con los HMM estándar, sino con otros HMM más complejos, como los modelos de Markov jerárquicos [202].

Incluso recientemente, este tipo de modelos se han aplicado a soluciones comerciales para el reconocimiento automático del habla. Gemello *et. al.* demostraron que los modelos híbridos no solamente obtienen mejores tasas de reconocimiento que los HMM clásicos y son más precisos a la hora de realizar la decodificación fonética, sino que además son compatibles e integrables en arquitecturas paralelas de alto rendimiento [84].

Aunque muchos modelos híbridos HMM/RNA en la literatura están enfocados al dominio del reconocimiento del habla, estos sistemas son igualmente aplicables a otros problemas de naturaleza secuencial. Rynkiewicz *et. al.* emplearon un HMM entrenado mediante algoritmos de máxima expectación junto a una red de neuronas para predecir una serie temporal generada mediante láseres, obteniendo una mejor segmentación de los datos y una precisión muy notable [212].

España-Boquera *et. al.* usaron una combinación de HMM con RNA para reconocimiento de texto manuscrito, obteniendo resultados muy satisfactorios, incluso en configuraciones de tiempo real [75]. También en el área del reconocimiento de texto, Marukatat *et. al.* entrenaron conjuntos de redes de neuronas para obtener las probabilidades de emisión de un HMM híbrido con la finalidad de reconocer frases manuscritas obtenidas de la base de datos UNIPEN [159].

Estos modelos también han demostrado ser útiles en el área del reconocimiento de señales biomédicas. Recientemente, Lazli *et. al.* han presentado un sistema híbrido, usando un perceptrón multicapa entrenado para el diagnóstico biomédico [144]. En dicho trabajo compararon el rendimiento de tres sistemas diferentes: un HMM discreto, un sistema híbrido HMM/RNA y redes de neuronas de base radial. Los autores concluyeron que los modelos más precisos y potentes para esta

tarea son los sistemas híbridos

Pero los sistemas HMM/RNA no son los únicos tipos de modelo híbrido presentes en la literatura. Otro enfoque más reciente que ha ido ganando popularidad es la combinación de modelos ocultos de Markov con máquinas de soporte de vectores. Al igual que pasa con los enfoques HMM/RNA, los modelos HMM/SVM se han empleado mayoritariamente en el dominio del reconocimiento automático del habla.

Ganapathiraju *et. al.* propusieron un sistema híbrido HMM/SVM que mapeaba las salidas del SVM con probabilidades a posteriori mediante el uso de una función sigmoide para el reconocimiento de dígitos manuscritos [82]. El sistema obtuvo una mejora significativa (más del 10 %) respecto a los HMM, requiriendo solamente una quinta parte de los datos de entrenamiento. El uso de la función sigmoide para la estimación de la probabilidad a posteriori es un recurso popular debido a su sencillez y rendimiento [218].

Rahbar *et. al.* incluyeron un modelo HMM/SVM como parte de un sistema de reconocimiento de palabras. El sistema propuesto incluía dos niveles: por un lado el primer nivel empleaba un algoritmo de agrupamiento para segmentar la señal de audio, y por otro el modelo HMM/SVM se encargaba de clasificar las palabras [198].

Otro dominio donde se han aplicado estos modelos híbridos ha sido la predicción de series temporales. Ghalwash *et. al.* propusieron emplear un modelo HMM/SVM en el ámbito biomédico para predecir series de tiempo multivariadas [85]. Los autores obtuvieron resultados muy prometedores, que mejoran la mejor tasa de clasificación que hasta entonces se había obtenido en su dominio, y requiriendo para ello tan sólo un 40 % de los datos de entrenamiento.

Aunque menos comunes, también existen trabajos que plantean sistemas híbridos con otro tipo de clasificadores. Las redes bayesianas han demostrado ofrecer buenos resultados cuando se combinan con HMM para reconocer fonemas [158, 157]. Aunque estos modelos no demostraron mejorar a los HMM clásicos, sí probaron ser sistemas muy flexibles y adaptables al dominio del reconocimiento del habla [157].

En lo referente a nuestro dominio, también se han presentado sistemas híbridos enfocados al reconocimiento de actividades humanas [147]. Una propuesta presentada por Lester *et. al.* planteaba un sistema híbrido que combinaba dos técnicas. Por un lado, emplearon una versión modificada del algoritmo AdaBoost [245] que seleccionara automáticamente los mejores atributos y generara conjuntos de modelos discriminativos para reconocer las actividades. Por otro, las estimaciones de los clasificadores eran empleadas para calcular las probabilidades a posteriori de las clases, y dichas probabilidades eran a su vez utilizadas como entradas al HMM. Los modelos discriminativos eran específicamente calibrados en la tarea de distinguir las actividades unas de otras, mientras que la capa del

HMM aseguraba las propiedades temporales del sistema y permitía realizar un seguimiento continuo de las actividades.

En esta tesis doctoral presentamos dos propuestas de modelos híbridos para el reconocimiento de actividades mediante redes de sensores inalámbricos. Evaluamos y comparamos el rendimiento de los modelos haciendo uso de diferentes conjuntos de datos obtenidos de entornos reales: tres conjuntos generados por Kasteren *et. al.*, los cuales han sido ampliamente usados en la literatura [128], y los conjuntos de datos obtenidos por nuestros entornos instrumentados. Este tipo de modelos se han aplicado para el reconocimiento de las actividades humanas utilizando dispositivos llevables, pero no se ha encontrado ningún estudio al respecto en el que haga uso de nuestra tecnología.

4.3.1. Arquitectura Híbrida para el Reconocimiento de AVDs

En esta tesis doctoral postulamos que la combinación de un modelo generativo y un modelo discriminativo mediante una arquitectura híbrida da lugar a mejores resultados que los obtenidos por dichos modelos empleados de forma individual, dentro del dominio del reconocimiento de actividades de la vida diaria.

Los modelos híbridos que proponemos beben de las bases de los sistemas propuestos por Bourlard para el reconocimiento del habla [26]. Nuestros modelos constan de dos componentes: 1) un clasificador estático que recibe la información de los sensores como entrada y emite como salida la probabilidad de pertenencia a cada clase, y 2) un modelo oculto de Markov que interpreta la salida del clasificador como la probabilidad de emisión de las observaciones y modela las relaciones temporales de las actividades del usuario.

Cada uno de esos componentes deben ser integrados en un único sistema, y tal y como se ha indicado anteriormente, nosotros proponemos combinar los HMM con dos modelos discriminativos que han demostrado funcionar muy bien en el dominio: las redes de neuronas artificiales y las máquinas de soporte de vectores.

Redes de Neuronas Artificiales

Formalmente, las RNA son un paradigma de procesamiento y aprendizaje inspirado en la forma en que funciona el sistema nervioso [97]. Consiste en un sistema constituido por una serie de neuronas interconectadas, que transforman un conjunto de variables de entrada en una variable de salida.

En un sentido más analítico, el modelo general de una red de neuronas artificial contendrá al menos los siguientes elementos:

- Un conjunto de entradas $x_{j,t}$, las cuales definen lo que se conoce como capa de entrada de la red.

- Un conjunto de pesos w_{ij} , cada uno de los cuales define la intensidad de la interacción entre una neurona presináptica j (que envía información) y una neurona postsináptica i (que recibe la información).
- Una regla de propagación $f_i[x_{j,t}, w_{ij}]$ que proporcionará el valor del potencial postináptico $p_{i,t} = f_i[x_{j,t}, w_{ij}]$.
- Una función de activación $L_i[a_{i,t-1}, p_{i,t}]$ que determina el estado de activación actual de la neurona i , que depende de su estado de activación anterior $a_{i,t-1}$ y de su potencial postsináptico actual.
- Una función de salida $S_i[a_{i,t}]$, que determinará la salida actual $y_i(t) = S_i[a_{i,t}]$, y que da lugar a lo que se conoce como capa de salida de la red.

Las neuronas se organizan en diferentes niveles o capas dentro de la red, a través de las cuales se transmite la información. Existen tres tipos de estas capas: la capa de entrada, que recibe los datos de entrada; las capas ocultas, donde se realiza el procesamiento; y la capa de salida, que ofrece la salida del modelo.

El proceso de aprendizaje de estos modelos suele estar basado en el algoritmo de retropropagación [100] (del inglés *Backpropagation*). Este algoritmo tiene por objeto de minimizar un error (normalmente definido de forma cuadrática) mediante la técnica del descenso de gradiente. Por ello, su funcionamiento se basa en el cálculo de las derivadas parciales del error del modelo con respecto a los valores obtenidos como salida por la red de neuronas y el valor real.

Este tipo de modelos pueden tratar con problemas no lineales siempre y cuando la función de activación de la red de neuronas sea no lineal (i.e. función sigmoideal o tangente hiperbólica). Los fundamentos y descripciones más detalladas acerca de estos modelos pueden encontrarse en [98]

Máquinas de Vectores de Soporte

Las SVM son un conjunto de algoritmos de aprendizaje supervisado que tratan de componer un clasificador óptimo lineal basándose en la idea de minimización del riesgo estructural (*Structural Risk Minimization* o SRM en inglés) [35].

Generalmente, los clasificadores están optimizados siguiendo algún criterio de minimización del riesgo empírico. La minimización del riesgo empírico, R_{emp} , es la base de diversas técnicas cuyo objetivo es encontrar un modelo que minimice el riesgo:

$$R_{emp}(\alpha) = \frac{1}{2l} \sum_{i=1}^l |y_i - f(x_i, \alpha)| \quad (4.5)$$

donde α define un conjunto de parámetros, y_i es la salida del sistema y x_i la entrada. Sin embargo, minimizar R_{emp} no necesariamente supone encontrar el

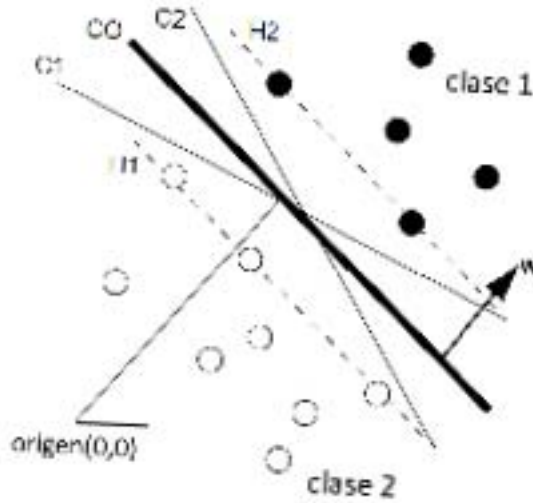


Figura 4.3: Ejemplos de hiperplanos para clasificación binaria.

clasificador óptimo. En el ejemplo mostrado en la Figura 4.3 aparece un problema de clasificación binario y las correspondientes regiones de decisión definidas por hiperplanos. Todos los hiperplanos ($C0$, $C1$ y $C2$) consiguen una clasificación perfecta, y por lo tanto cero valor de riesgo. Sin embargo, se puede apreciar cómo $C0$ es el hiperplano óptimo porque maximiza la distancia entre los límites $H1$ y $H2$, y por tanto obtiene la mejor generalización [242]. Este modo de aprendizaje es un ejemplo de minimización del riesgo estructural, donde el objetivo no es minimizar el riesgo empírico, sino generar un clasificador que maximice el margen geométrico entre muestras de diferentes clases [242].

Las SVM mapean los puntos de entrada con un espacio de características de una dimensión mayor, para después encontrar el hiperplano que los separe y maximice el margen entre las clases. Por tanto, el proceso de optimización para estos modelos se basa en encontrar los parámetros de un hiperplano que garantice una buena clasificación de los datos de entrenamiento y la maximización de los límites (la distancia entre $H1$ y $H2$ en la Figura 4.3). Los puntos que conforman el hiperplano se denotan mediante la función 4.6.

$$\vec{w} \cdot \vec{x} + b = 0 \quad (4.6)$$

Donde w es un vector normal al hiperplano y b es la desviación del hiperplano desde el origen. Considerando que existen N ejemplos de entrenamiento ($i = 1, \dots, N$) con clase y_i (con valores $+1$ o -1), se satisfacen las siguientes restricciones:

$$\begin{aligned} \vec{x}_i \cdot \vec{w} + b &\geq +1 & \text{para } y_i = +1 \\ \vec{x}_i \cdot \vec{w} + b &\leq -1 & \text{para } y_i = -1 \end{aligned} \quad (4.7)$$

Empleando el método de los multiplicadores de Lagrange, el problema de maximización de los límites se convierte en un problema de optimización cuadrática, y da como resultado un conjunto de ejemplos de entrenamiento (los llamados vectores de soporte), que son los que definen el hiperplano. Una vez que la SVM ha sido entrenada, la clasificación puede llevarse a cabo mediante la ecuación 4.8.

$$y(t) = \vec{x}(t) \cdot \vec{w} + b \quad \text{donde} \quad \vec{w} = \sum_{i=1}^{N_S} \alpha_i y_i \vec{s}_i \quad (4.8)$$

donde N_S es el número de vectores de soporte \vec{s}_i , y y_i su etiqueta de clase.

Debido a que la mayoría de los problemas del mundo real no son linealmente separables (incluyendo el reconocimiento de AVD), se necesita extender la formalización del problema introduciendo un término de regularización que depende de variables de holgura, tal y como se recoge en la formulación 4.9.

$$\begin{aligned} \vec{x}_i \cdot \vec{w} + b &\geq +1 - \xi_i & \text{para } y_i = +1 \\ \vec{x}_i \cdot \vec{w} + b &\leq -1 + \xi_i & \text{para } y_i = -1 \\ \xi_i &\geq 0 \end{aligned} \quad (4.9)$$

El número de errores en el entrenamiento puede ser definido como $\sum_i \xi_i$.

$$K(\vec{w}, \vec{x}) = \phi(\vec{w}) \cdot \phi(\vec{x}) \quad (4.10)$$

La característica de las SVM es su habilidad para manejar problemas no lineales mediante el mapeo del espacio de características a un espacio de dimensión mayor, donde el problema sí que sea linealmente separable. Esto se realiza mediante una representación por medio de funciones Kernel. Una función núcleo o kernel es un producto interno en el espacio de características, que tiene su equivalente en el espacio de entrada. Debido a la propiedad de las funciones kernel ($K(\vec{w}, \vec{x}) = \phi(\vec{w}) \cdot \phi(\vec{x})$), la información se puede proyectar a un espacio de características de mayor dimensión, y por tanto el clasificador final pasa a ser como se indica en la ecuación 4.11. Una descripción más detallada y una formalización más completa de este paradigma se puede consultar en [54].

$$y(t) = \sum_{i=1}^{N_S} \alpha_i y_i K(\vec{s}_i \cdot \vec{x}(t)) + b \quad (4.11)$$

Sistema HMM Híbrido

Para problemas de reconocimiento como el que se plantea en esta tesis doctoral, la función del clasificador que está integrado en el modelo híbrido debe ser el de aproximar probabilidades. Concretamente, dadas las funciones que definen un HMM, nosotros queremos estimar la probabilidad de emisión $p(\vec{x}|\mathbf{y})$ (ver Función 4.2), es decir, la probabilidad de observar unos valores de los sensores dado un hipotético estado del HMM (el cual se corresponde con la actividad de la persona). Sin embargo, los HMM están basados en formalismos muy estrictos que dificultan que sea posible llevar a cabo modificaciones sobre el modelo sin perder la base teórica de sus algoritmos de aprendizaje e inferencia. No obstante, si se consigue que los clasificadores estimen probabilidades relacionadas con las probabilidades de emisión del modelo, estos pueden ser muy fácilmente combinados con el HMM.

En lo relativo a las redes de neuronas, éstas pueden ser entrenadas para estimar la probabilidad a posteriori $p(\mathbf{y}|\vec{x})$, es decir, la probabilidad para una clase \mathbf{y}_t dado el vector de características \vec{x} . En estudios previos se ha demostrado que una RNA entrenada para llevar a cabo clasificación supervisada es un estimador de probabilidad a posteriori, si se asocia cada neurona de salida a un estado específico del HMM [21]. Para ello, el mecanismo más común para obtener dicha distribución en cada estado $\mathbf{y} \in \{1, \dots, Q\}$, es mediante el empleo de un enfoque probabilístico que considera la función de activación softmax en cada uno de los nodos de salida. La función a aplicar es por tanto:

$$p(\mathbf{y}|\vec{x}) = \frac{\exp(s_{\mathbf{y}})}{\sum_{i \in Q} \exp(s_i)} \quad (4.12)$$

Donde $s_{\mathbf{y}}$ es el \mathbf{y} -ésimo valor de salida antes de aplicar la función de activación.

Por ello, la estimación obtenida por la RNA, cuando se introduce como entrada \vec{x} , será un cálculo de la probabilidad a posteriori $p(\mathbf{y}|\vec{x})$ sobre la clase \mathbf{y} [194, 205, 76]. En la Figura 4.4 se muestra un diseño general de nuestra propuesta en esta situación, cuando el modelo discriminativo empleado es una RNA.

Este proceso es válido para cualquiera de las arquitecturas conexionistas clásicas, tales como el perceptrón multicapa, siempre y cuando se den dos condiciones:

1. Debe existir un número suficientemente alto de ejemplos de entrenamiento para aprender una buena aproximación entre las capas de entrada y salida.
2. El sistema debe ser entrenado para encontrar un error mínimo global (por ejemplo, empleando como criterio el error cuadrático medio o la entropía relativa).

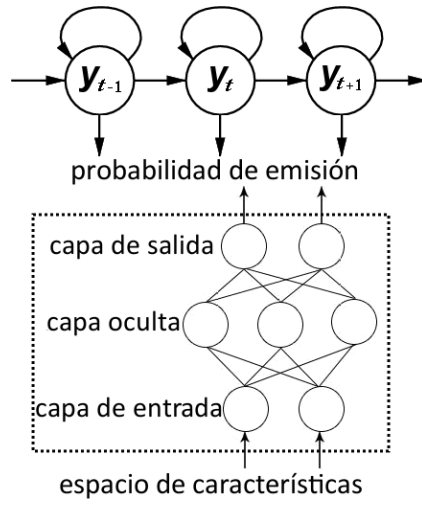


Figura 4.4: Estructura del modelo híbrido HMM/RNA empleado.

En lo relativo a las máquinas de soporte vectorial, las salidas de estos algoritmos representan una medida de distancia, por lo tanto, para poder ser combinados con un HMM, deben transformarse en probabilidades a posteriori.

Una forma eficiente de estimar estas probabilidades es mediante el empleo de una función sigmoide en cada una de las salidas de la SVM [190]. Según este método la probabilidad a posteriori para la clase y , dado el vector de características \vec{x} , está definida mediante la función $p(y|\vec{x})$ descrita en la ecuación 4.13.

$$p(y|\vec{x}) = \frac{1}{1 + A \exp(-s_y + B)} \quad (4.13)$$

donde s_y simboliza la salida de la SVM para la clase y , y los parámetros A y B son calculados tal y como se propone en [95].

Una vez se ha obtenido la probabilidad a posteriori a través del clasificador, las probabilidades de emisión pueden ser estimadas usando la Fórmula de Bayes (ver ecuación 4.14).

$$p(\vec{x}|y) = \frac{p(y|\vec{x})p(\vec{x})}{p(y)} \quad (4.14)$$

Las probabilidades a priori de cada clase $p(y)$ pueden ser obtenidas de manera frecuentista usando los datos de entrenamiento. Por tanto, las probabilidades $p(y|\vec{x})/p(y)$ puede ser usadas como probabilidades de emisión del HMM, ya que el factor de escala $p(\vec{x})$ es constante para todas las clases [26].

El modo de aplicar esta transformación a nuestra secuencia de vectores binarios de los sensores es mediante un proceso como el que aparece representado en la Figura 4.5. Para cada instante t , existe un vector de características \vec{x}_t obtenido

de la red de sensores, que sirve como entrada del clasificador estático de modelo híbrido (representando mediante el conjunto de parámetros Θ). El clasificador emite para cada instante una probabilidad a posteriori, que es transformada tal y como se ha descrito antes, dando lugar a la probabilidad de emisión b_y para la clase y .

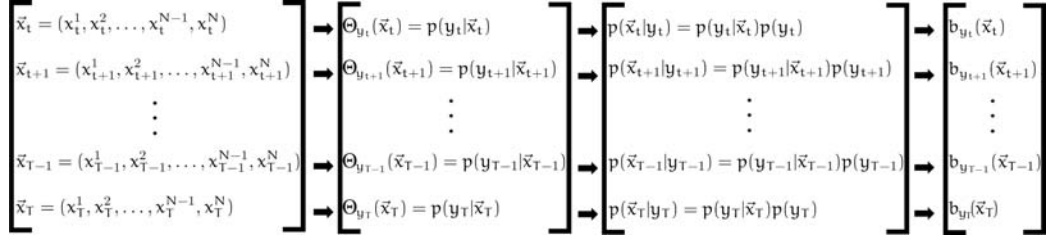


Figura 4.5: Esquema genérico para el procesamiento de nuestros datos en un modelo HMM híbrido.

El entrenamiento del modelo híbrido se realiza mediante un algoritmo iterativo de máxima expectación, como el propuesto en [76], donde se alternan el entrenamiento supervisado del clasificador y un algoritmo de inferencia. El proceso completo se describe como sigue:

1. De entre todos los datos etiquetados, se seleccionan los subconjuntos de datos que van servir como ejemplos de entrenamiento y test para el clasificador. Siguiendo la notación del Apartado 3.3, cada instancia de datos se compone del vector de valores lógicos que representa la activación de los sensores y la clase que representa la actividad. Esto es, la instancia que utiliza nuestro sistema para el instante t está representada por $I_t = (x_t^1, x_t^2, \dots, x_t^{N-1}, x_t^N, y_t)$.
2. Se asignan valores iniciales a los parámetros $\theta = \{\pi, A\}$ del HMM, distintos de cero.
3. Se lleva a cabo el entrenamiento de la RNA (o de la SVM) usando el conjunto de entrenamiento.
4. Mediante el algoritmo de *Viterbi* y las estimaciones del clasificador se estima la secuencia de estados más probable. Para dicha estimación se emplean las probabilidades de transición a priori obtenidas a partir de la frecuencia relativa de cada clase en los ejemplos de entrenamiento. Las partes del algoritmo de *Viterbi* en donde se requiere estimar la probabilidad de emisión para un estado, se sustituyen por el proceso de estimación mediante Bayes descrito anteriormente.

Tabla 4.1: Tipos de actividades y porcentaje de instancias por clase para cada uno de los conjuntos de datos.

| Actividad | KasterenA | KasterenB | KasterenC | OrdonezA | OrdonezC |
|---------------------|-----------|-----------|-----------|----------|----------|
| Salir | 49.74 % | 54.36 % | 46.27 % | 8.32 % | 17.41 % |
| Usar el inodoro | 0.65 % | 0.27 % | 0.62 % | 0.76 % | 0.55 % |
| Ducha | 0.7 % | 0.6 % | 0.6 % | 0.54 % | 0.24 % |
| Dormir | 33.42 % | 33.53 % | 28.46 % | 39.1 % | 35.58 % |
| Desayunar | 0.23 % | 0.52 % | 0.62 % | 0.63 % | 1.02 % |
| Cenar | 1.0 % | 0.42 % | 1.26 % | 0 % | 0.38 % |
| Beber | 0.1 % | 0.07 | 0.11 % | 0 % | 0 % |
| Sin etiquetar | 14.12 % | 10.12 % | 21.97 % | 5.61 % | 11.73 % |
| Almorzar | 0 % | 0 % | 0 % | 1.59 % | 1.30 % |
| Tomar un refrigerio | 0 % | 0 % | 0 % | 0.05 % | 1.33 % |
| Ocio/TV | 0 % | 0 % | 0 % | 42.7 % | 28.98 % |
| Aseo | 0 % | 0 % | 0 % | 0.73 % | 1.42 % |

5. Se vuelve al paso 3 hasta que el método converge, esto es, hasta que la diferencia entre dos iteraciones consecutivas no supera un cierto límite.

4.3.2. Evaluación

En este apartado presentamos los experimentos que hemos llevado a cabo para evaluar nuestra propuesta. Describimos los conjuntos de datos, la configuración de los experimentos, el modo elegido de representar los datos y los resultados de las pruebas. El primer experimento está enfocado a determinar el mejor tamaño de la ventana deslizante para los modelos discriminativos. En este primer experimento se incluyen tanto los algoritmos discriminativos que después se integran en el modelo híbrido, como varios de los algoritmos de clasificación supervisada más populares en la literatura, con la finalidad de ilustrar el rendimiento de nuestra propuesta. En el segundo experimento comparamos el rendimiento de nuestro sistema con el resto de modelos que se han descrito en este capítulo de la tesis.

Conjuntos de Datos

Los experimentos se han llevado a cabo usando tanto los conjuntos de datos generados por nosotros como varios conjuntos de datos libremente disponibles para la comunidad investigadora.

Nuestros conjuntos de datos son aquellos obtenidos a través de los entornos de monitorización descritos en el Apartado 3.1.3. Sin embargo, debemos hacer notar que nuestro conjunto de datos ‘OrdonezB’ no pudo ser empleado para esta parte del trabajo, debido a una mala anotación de las actividades por parte del usuario, lo que lo inutiliza para ser empleado en tareas de aprendizaje supervisado.

Además de nuestros propios datos hemos hecho uso de los conjuntos de datos libres que crearon Kasteren *et. al.* [125]. Estos conjuntos de datos hacen uso del mismo espacio de características que nuestros datos, y es por ello que nos sirven para aumentar la evaluación de nuestra propuesta.

En la Tabla 4.1 se muestra qué actividades fueron etiquetadas y su proporción en cada uno de los conjuntos. Se puede apreciar cómo el porcentaje de instancias para cada clase varía en función del conjunto, y cómo determinadas actividades (como por ejemplo ‘Beber’) ni siquiera aparecen según qué conjunto de datos.

Representación de los Datos

Los flujos de datos generados por la red de sensores pueden ser usados directamente tal y como están representados en la Figura 3.7 o ser preprocesados empleando diferentes tipos de representación.

Con objeto de aumentar el espacio de características y para obtener una mejor evaluación, en esta tesis doctoral hemos llevado a cabo los experimentos empleando tres tipos diferentes de representación de datos, previamente validados en nuestro dominio [128]. Estas tres formas de representación son:

- *Raw*: La representación de los datos en bruto o *raw* muestra los datos de los sensores de forma directa, del mismo modo que son recibidos del entorno de monitorización. El valor del intervalo es 1 si el sensor está activo y 0 en caso contrario (ver Figura 4.6(a)).
- *ChangePoint*: La representación de punto de cambio o *change point* indica el momento en el que el sensor binario cambia su estado. Esto es, el valor en el intervalo es 1 si el sensor ha cambiado su estado de cierto a falso o viceversa, y es 0 en caso contrario (ver Figura 4.6(b)).
- *LastSensor*: La representación de última activación o *last sensor* muestra qué sensor se ha activado en último lugar. Según esta representación en cada intervalo de tiempo solamente puede estar representado uno de los sensores como activo. El último sensor que haya cambiado de estado seguirá mostrándose como activo mientras otro sensor no cambie su valor (ver Figura 4.6(c)).

El modo de representación *raw* es el más intuitivo debido a que el valor x_t^i definido para el sensor i en el instante t , está directamente relacionado con la salida

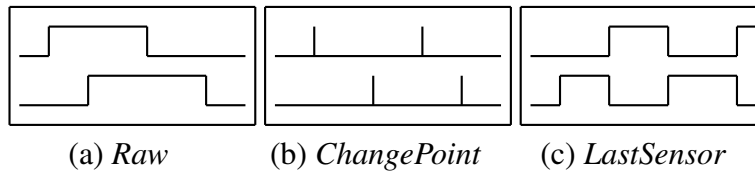


Figura 4.6: Tipos de representación de los datos.

binaria del sensor. Sin embargo, este modo de representación plantea algunos problemas cuando se aplica a datos obtenidos de un entorno monitorizado real. Esto es porque hay sensores, como por ejemplo los sensores magnéticos de las puertas, que no cambian su estado de forma automática, sino que requieren de algún movimiento mecánico para ello. En la práctica, esto significa que si el usuario no cierra bien un armario, en el sistema aparecerá como que este armario está siendo usado constantemente. Desde el punto de vista del reconocimiento de patrones, implica que el valor discriminativo de ese sensor se reduce drásticamente.

El modo de representación *changepoint* soluciona dicho problema de activación continua de los sensores. Además proporciona información útil debido a que el cambio de estado de algunos sensores coincide consistentemente con el punto inicial o final de una actividad. Pero el empleo de este tipo de representación implica que el sistema tiene largos periodos de tiempo en donde no aparecen activaciones de los sensores, principalmente para ciertas actividades de larga duración, como por ejemplo dormir. Esto da lugar a que la observación de ningún sensor activo sea muy ambigua cuando se usa esta representación.

El problema de la ambigüedad se soluciona con el modo de representación *lastsensor*. La motivación de este tipo de representación es que el último sensor que se activó es muy indicativo de la localización de la persona en la vivienda. Y dependiendo de la disposición de la vivienda, esta representación puede favorecer el reconocimiento de las actividades diarias.

En estudios previos, se ha visto cómo las mejores tasas de reconocimiento suelen estar asociadas a la representación de los datos mediante *changepoint* o *lastsensor* [132]. La combinación de estos modos de representación puede dar lugar a un ligero aumento en la precisión de los modelos, pero las diferencias no son estadísticamente significativas.

Configuración de la Experimentación

La división de los datos en entrenamiento y test la llevamos a cabo haciendo un tipo de validación cruzada ‘dejando uno fuera’ (*Leave-one-out cross-validation* o LOOCV en inglés). Este tipo de validación cruzada implica separar los datos de forma que para cada iteración exista solamente una muestra para los datos de prueba y el resto lo conformen los datos de entrenamiento. En nuestro caso

dejamos fuera un día entero de datos de los sensores, el cual será usado para prueba, y los días restantes se emplean para entrenar los modelos. Dicho proceso se repite para todas las iteraciones, empleando en cada iteración una submuestra diferente y obteniendo así un resultado promedio.

Como se puede apreciar en la Tabla 4.1, los conjuntos sufren de un problema de desbalanceo de datos, debido a la naturaleza del dominio. Este tipo de problema aparece cuando una o algunas clases están representadas por muchos ejemplos en el conjunto, mientras que del resto de las clases existen pocas instancias [121]. Esta desproporción en la representatividad de las clases da lugar a una disminución en la efectividad de la clasificación sobre las clases menos representadas.

Un clasificador trivial que no fuese capaz de generalizar, y que determinase que todos los ejemplos pertenecen a la clase mayoritaria podría obtener una precisión elevada. Cuando se debe aprender sobre datos que se encuentren muy desbalanceados, el acierto promedio en la clasificación no se considera una medida de precisión adecuada.

En nuestro caso, las clases menos frecuentes son de gran interés para nuestro problema, y por ello empleamos como métricas el Valor Predictivo Positivo (PPV) (ver ecuación 4.15), la Razón de Verdaderos Positivos (VPR) (ver ecuación 4.16) y una métrica que combina ambos valores, la medida F o *F-Measure* (ver ecuación 4.17). Estas métricas pueden ser calculadas usando una matriz de confusión como la que aparece en la Tabla 4.2.

$$\text{PPV} = \frac{1}{N} \sum_{i=1}^N \frac{VP_i}{TP_i} \quad (4.15)$$

$$\text{VPR} = \frac{1}{N} \sum_{i=1}^N \frac{VP_i}{TR_i} \quad (4.16)$$

$$\text{F-Measure} = \frac{2 \cdot \text{PPV} \cdot \text{VPR}}{\text{PPV} + \text{VPR}} \quad (4.17)$$

Las filas de la tabla representan las instancias en la clase real, tal y como han sido etiquetadas por los usuarios, mientras que las columnas representan el número de predicciones de cada clase. La diagonal de la matriz contiene los verdaderos positivos (VP), mientras que la suma de una fila nos da el número de instancias reales para esa clase (TR) y la suma de una columna el total de predicciones para la clase correspondiente (TP). Nosotros estimamos el VPR y el PPV para cada clase de forma independiente, y después calculamos el valor promedio para el conjunto. Con las métricas empleadas se considera la clasificación de cada clase igualmente importante.

Para estimar si la diferencia entre dos casos *A* y *B* es estadísticamente significativa empleamos dos test estadísticos diferentes. Por un lado, empleamos una prueba *t* de Student de una cola para muestras pareadas, con nivel de significación

Tabla 4.2: Matriz de confusión que muestra los verdadores positivos (VP), el total de las clases reales (TR) y el total de las clases predichas (TP) para cada clase.

| Real | Predicho | | | |
|------|-----------------|-----------------|-----------------|---------|
| | 1 | 2 | 3 | |
| 1 | VP_1 | ϵ_{12} | ϵ_{13} | TR_1 |
| 2 | ϵ_{21} | VP_2 | ϵ_{23} | TR_2 |
| 3 | ϵ_{31} | ϵ_{32} | VP_3 | TR_3 |
| | TP_1 | TP_2 | TP_3 | $Total$ |

de 0.05. Esto implica que el rendimiento en el caso A para una iteración en particular de la validación cruzada (un día de datos) se compara con el rendimiento en el caso B para ese mismo día de datos de dicha validación cruzada. La comparación se realiza de este modo porque existen grandes diferencias en el tipo de actividades que aparecen en los datos, dependiendo del día del que se trate. Es por ello, que para obtener una comparación justa, el rendimiento de los modelos para cada día debe calcularse de forma pareada.

Sin embargo, debido a que se ha demostrado que la prueba t de Student en algunos casos tiene alta probabilidad de cometer errores tipo I (también denominado error de tipo alfa) cuando ésta se aplica a una validación cruzada del tipo ‘dejando uno fuera’ [66], hemos complementado nuestro test de significancia con la prueba de los rangos con signo de Wilcoxon. Un error de tipo I se comete cuando la hipótesis nula es verdadera y como consecuencia del contraste se rechaza, es decir, es lo que se conoce como un falso positivo. La prueba de los rangos con signo de Wilcoxon es un test no paramétrico empleado para comparar la media de dos muestras relacionadas y determinar si existen diferencias entre ellas [65]. Es un test estadístico más seguro que la prueba t de Student debido a que no asume una distribución normal u homogeneidad en la varianza. Por tanto, puede usarse con los datos de precisión de los clasificadores, las tasas de error o sobre cualquier otra métrica de evaluación de los modelos. Resultados empíricos muestran cómo este test es particularmente adecuado cuando se comparan pares de clasificadores [65].

El modelo híbrido propuesto lo hemos comparado tanto con los modelos generativos secuenciales presentados en otros estudios (HMM) [128], como con algoritmos discriminativos que hacen uso de una ventana deslizante.

Un dato de especial interés en esta evaluación es poder comprobar si el hibridar dos modelos clasificadores supone una mejora con respecto al rendimiento de cualquiera de esos modelos de forma independiente. Es decir, en este traba-

jo de tesis doctoral nos interesa conocer si nuestra propuesta híbrida es capaz de mejorar tanto al HMM que controla la secuencialidad de las actividades como al modelo discriminativo que clasifica las observaciones de los sensores. Por ello, el modelo híbrido propuesto lo hemos comparado tanto con los modelos generativos secuenciales presentados en otros estudios (HMM) [128], como con algoritmos discriminativos que hacen uso de una ventana deslizante.

La elección mas inmediata respecto a los algoritmos discriminativos a introducir en la comparativa es elegir los modelos que hemos seleccionado para ser integrados en el sistema híbrido: las RNA y las SVM. De este modo, podremos evaluar cuál es la diferencia en rendimiento para estos algoritmos en función de si se emplean de forma independiente o de si se combinan con un HMM.

Además, de cara a obtener una mayor validación de nuestra propuesta, también se incluyen en la comparativa los modelos que hemos seleccionado para ser integrados en el sistema híbrido: las RNA y las SVM. De este modo, también podremos evaluar cuál es la diferencia en rendimiento para estos algoritmos en función de si se emplean de forma independiente o de si se combinan con un HMM.

Además, con el objeto de obtener una mayor validación de nuestra propuesta, se han introducido otros algoritmos discriminativos, tomando como referencia el trabajo de Bao *et. al.*, donde se realizó un experimento de similares características, también en el ámbito del reconocimiento de actividades [11]. De cara a introducir en nuestra comparativa la mayor representatividad de los diferentes tipos de aprendizaje, los algoritmos discriminativos que se han seleccionado han sido:

1. Aprendizaje basado en reglas. El algoritmo que representa este tipo de aprendizaje es el algoritmo de clasificación *DecisionTable* [138]. Este es un algoritmo muy sencillo que divide los datos usando una matriz que está dividida en condiciones y acciones. Las condiciones definen los atributos de las instancias y las acciones definen las instancias etiquetadas del espacio de características. Para clasificar un nuevo ejemplo, el algoritmo busca una coincidencia exacta usando las condiciones. En función de cuantas instancias de la tabla resultan coincidentes, el algoritmo asigna la clase al nuevo ejemplo.
2. Aprendizaje basado instancias. El algoritmo que representa este tipo de aprendizaje es el algoritmo de clasificación Vecino Más Cercano, también conocido como $k - NN$ (*K-Nearest Neighbor*) [56]. Este algoritmo compara cada nuevo ejemplo con todos los observados previamente y le asigna la clase del ejemplo más similar. Para ello, utiliza la distancia Euclídea normalizada con la que detecta las instancias de entrenamiento más cercanas

a la que se quiere clasificar. Este algoritmo necesita establecer inicialmente el valor de k , que indica el número de instancias más cercanas que se van a tener en cuenta para clasificar un ejemplo dentro de una clase determinada. En este caso, se ha determinado de forma experimental el valor $k = 5$, por obtener con este valor los mejores resultados.

3. Aprendizaje basado en árboles. El algoritmo que representa este tipo de aprendizaje es el algoritmo de clasificación C4.5 [196]. Este algoritmo está basado en la creación de árboles de decisión a partir de una estrategia de divide y vencerás. Cada nodo del árbol está asociado con un conjunto de instancias a las que se les asignan ciertos pesos para tener en cuenta valores de atributos no conocidos.

El tipo de RNA que hemos empleado para llevar a cabo la tarea de clasificación es un perceptrón multicapa, entrenado mediante retropropagación y con una arquitectura de tres capas. Los nodos de la capa de entrada están definidos en función del espacio de características y los nodos de la capa de salida en función del número de actividades a reconocer. En lo relativo a la capa oculta, no existe una heurística que nos ayude a determinar cuál es el mejor número de nodos sin tener que entrenar varias veces el clasificador para estimar su error. Si elegimos para esta capa un número de nodos demasiado pequeño, obtendremos un alto error en la generalización y en el entrenamiento debido al sesgo estadístico. Si por el contrario definimos un número de nodos demasiado alto, podremos obtener un error bajo en el entrenamiento, pero todavía tendremos un error alto en la clasificación debido a la sobreadaptación del modelo y a la alta varianza [83]. Una regla bastante simple para la selección del número de nodos en la capa oculta es la enunciada por Blum, que propone que dicho valor sea el término medio entre el número de nodos en las capas de entrada y de salida [23]. Otra regla en este sentido la proponen Berry *et. al.*, los cuales indican que el número de nodos en la capa oculta nunca debe exceder el doble del número de nodos de la capa de entrada [18].

En lo relativo a las SVM, éstas se deben parametrizar mediante una función de kernel o núcleo. En nuestro caso, la función más popular [40] es también la que ha ofrecido mejores resultados: la función de base radial (*Radial Basis Function* o *RBF* en inglés). La función de kernel *RBF*, sobre dos muestras x y y , se define según la función

$$K(x, y) = \exp\left(-\frac{|x - y|^2}{2\sigma^2}\right) \quad (4.18)$$

donde $|x - y|^2$ es la distancia euclídea cuadrática entre los dos vectores de características.

Tamaño de la Ventana Deslizante

En este apartado presentamos las pruebas que hemos llevado a cabo para determinar el tamaño más adecuado de ventana para los modelos discriminativos. Los experimentos se han realizado empleando los tres tipos de representación de los datos detallados en el Apartado 4.3.2, y el resultado que se presenta es el valor promediado.

Tal y como se ha indicado anteriormente, para poder discretizar el flujo de datos se debe segmentar la información de los sensores en intervalos regulares, con una granularidad de tiempo Δt . En cada intervalo t se define una instancia de datos, que contiene los valores de los sensores para ese segmento de tiempo y donde la clase está representada por la actividad que ha realizado el usuario. Para este trabajo asumimos que a todos los intervalos de tiempo les corresponde una actividad, y que las actividades no las realiza el usuario de forma simultánea, de tal modo que cada instante de tiempo está asociado con una única actividad discreta.

Elegir la granularidad de tiempo Δt adecuada es importante para nuestra tarea de reconocimiento. Segmentar una señal usando intervalos demasiado pequeños añadiría ruido a los datos, y el empleo de intervalos demasiado grandes diluiría los detalles importantes de la señal.

Para elegir el valor Δt nos hemos basado en las contribuciones realizadas por Kasteren *et. al.* [132]. Tras varios experimentos, los autores concluyeron que la división del flujo de datos empleando intervalos de 60 segundos era la configuración que obtenía el mejor rendimiento de los modelos, a la vez que respetaba la representatividad de las actividades. El uso de intervalos de tiempo más grandes introducía un sesgo estadístico a los datos, puesto que provocaba que el solapamiento de varias actividades en un mismo intervalo fuese más común, y en última instancia que las actividades breves llegasen a ser irrelevantes para la tarea de clasificación. Tras el proceso de discretización se tienen 35486 intervalos para el conjunto ‘KasterenA’, 19968 intervalos para el conjunto ‘KasterenB’, 26236 intervalos para el conjunto ‘KasterenC’, 20160 intervalos para el conjunto ‘OrdonezA’ y 30240 para el conjunto ‘OrdonezC’.

Dicho intervalo nos sirve para determinar con qué frecuencia generamos una instancia del flujo temporal, pero igualmente el problema sigue siendo secuencial y, dado que en este trabajo vamos a aplicar algoritmos de clasificación bajo una ventana deslizante, nos interesa determinar cuál es el ancho de ventana más adecuado al problema que se quiere resolver. Es decir, dada una ventana como la que hemos detallado anteriormente $\langle \vec{x}_{t-w}, \vec{x}_{t-w+1}, \dots, \vec{x}_t \rangle$, buscamos determinar la longitud w que ofrezca mejores resultados en este problema. Esto es, el clasificador estimará la clase y_t no en función de un vector de valores binarios, sino de w vectores.

En la Figura 4.7 se pueden observar los valores de *F-Measure* para todos los algoritmos discriminativos propuestos en cada conjunto de datos. Se puede apreciar cómo ninguna longitud de ventana en concreto es capaz de obtener el mejor rendimiento de forma consistente. Para la mayoría de los modelos, un aumento en la longitud de la ventana implica una mejora en la clasificación, aunque la mejoría en muchos de los casos no llega a ser estadísticamente significativa. Incluso en algunos casos, como cuando se aplica el algoritmo basado en reglas con ventana deslizante de longitud 5, el rendimiento del modelo empeora (ver Figura 4.7(b)(c)).

Para determinar cuál es el tamaño de ventana más adecuado al problema planteado debemos considerar la Figura 4.8, donde aparece el rendimiento de los modelos de forma promediada para todos los conjuntos de datos. Lo más reseñable es que las diferencias entre usar una ventana de longitud 1 y una ventana de longitud 3 son estadísticamente significativas para gran parte de los modelos, excepto para el algoritmo basado en reglas y el algoritmo basado en instancias.

La explicación de que el algoritmo basado en reglas empeore de forma progresiva con el aumento de la ventana se debe a la simplicidad del mismo. Debido a que no es capaz de asignar pesos a los atributos, los considera todos igualmente importantes, y aumentar el espacio de características introduciendo activaciones de los sensores que pueden no estar relacionadas con la actividad a clasificar, tiene como consecuencia que se añade ruido para este modelo.

También, viendo la Figura 4.8, se aprecia que la diferencia entre usar una ventana de longitud 3 y una ventana de longitud 5 realmente no es significativa, por lo que es presumible que aumentar aún más el espacio de características no supone mejorar el rendimiento de los modelos.

A la vista de los resultados de este experimento, decidimos que la longitud de ventana adecuada a nuestro problema es 3, lo cual se traduce en que los modelos discriminativos tendrán en consideración las activaciones de los sensores durante los tres minutos anteriores al instante cuya actividad deben clasificar. La idea detrás de usar una ventana deslizante y considerar también valores anteriores en el flujo de datos, es proporcionar a los clasificadores datos de los sensores previos al momento en el que se desarrolla la actividad. Esta información es útil porque algunas actividades están estrechamente correladas con algún sensor, y puede darse el caso de que la segmentación de los datos suponga en algunos casos (por ejemplo cuando el inicio de la actividad está muy cercana al principio del intervalo) dividir la tupla sensor-actividad en instancias diferentes.

Comparativa de los modelos

En este apartado se muestra el rendimiento global para cada uno de los modelos introducidos en la comparativa. Los diferentes modelos se comparan en

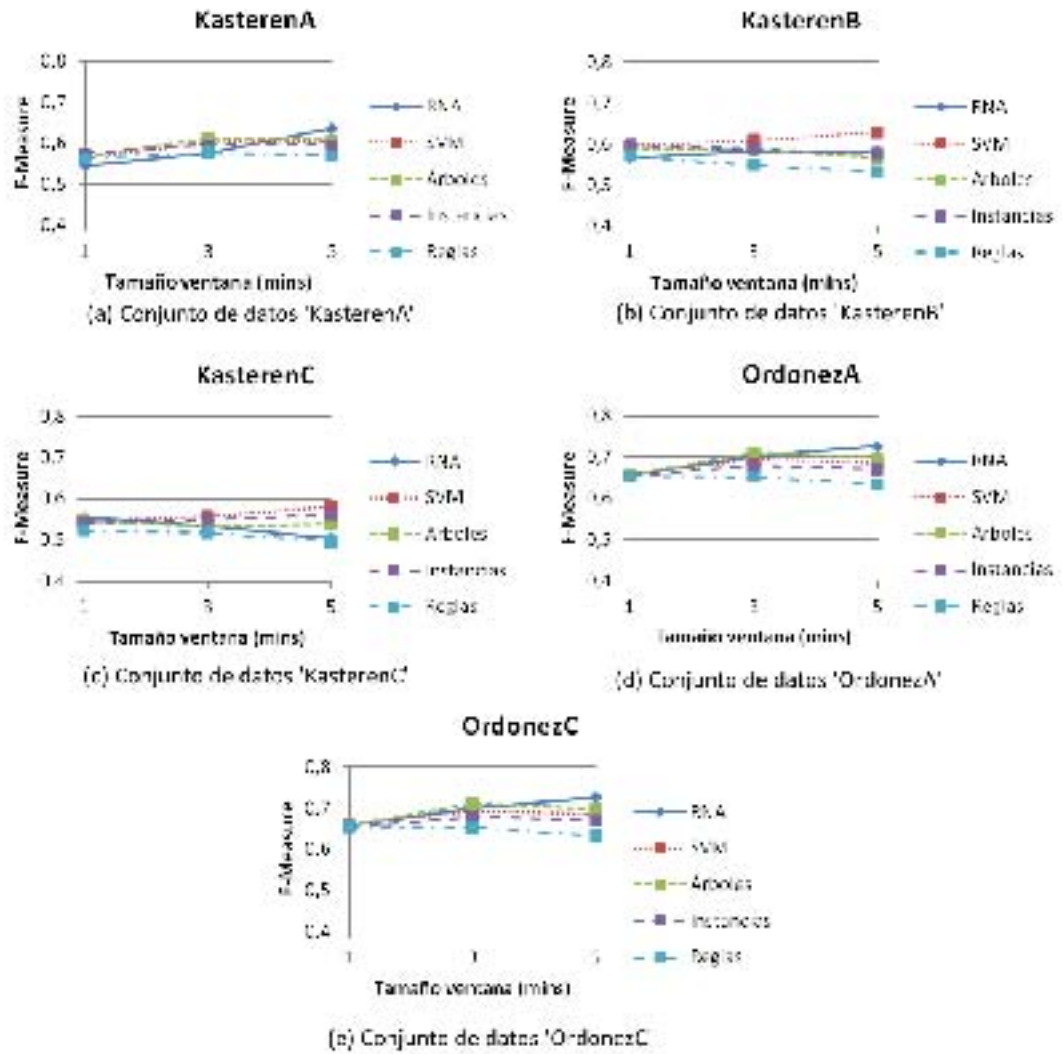


Figura 4.7: Valores de F -Measure para los modelos discriminativos en cada conjunto de datos, usando diferentes longitudes de ventana deslizante.

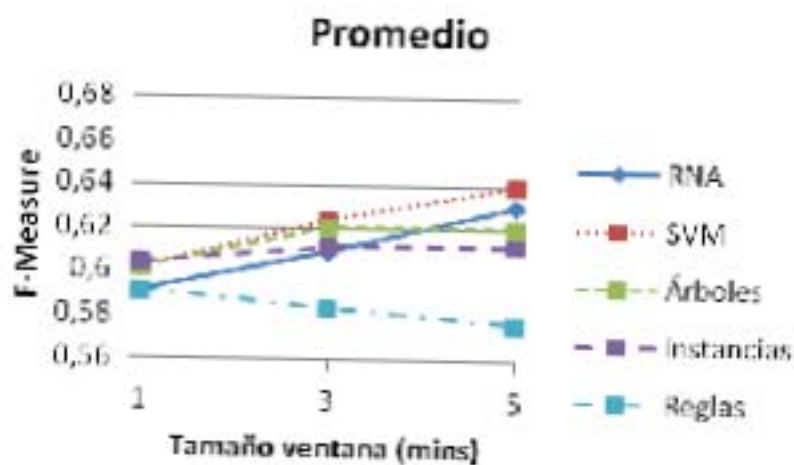


Figura 4.8: Valor promediado de *F-Measure* para todos los conjuntos de datos empleados usando los modelos discriminativos y diferentes longitudes de ventana deslizando.

términos de *F-Measure* empleando los tres tipos de representación de los datos detallados en el Apartado 4.3.2.

Los resultados para los conjuntos de datos ‘KasterenA’, ‘KasterenB’, ‘KasterenC’, ‘OrdonezA’ y ‘OrdonezC’ se muestran en las Tablas 4.3, 4.4, 4.5, 4.6 y 4.7 respectivamente. Para cada tabla, las columnas muestran el rendimiento de cada uno de los modelos, mostrando porcentualmente tanto el valor de la métrica como su desviación estándar.

En lo referente a la precisión obtenida según el tipo de representación empleada, los valores para el modelo discriminativo son consistentes con los resultados en la literatura, ofreciendo un mejor rendimiento con las representaciones *ChangePoint* y *LastSensor*. Respecto a este punto, es interesante ver cómo los clasificadores ofrecen consistentemente el peor rendimiento con la representación *ChangePoint*. Esto es debido a su método de aprendizaje, ya que cuando se emplea este tipo de representación existen actividades (‘Dormir’, ‘Salir’, ‘Sin etiquetar’) que comparten los valores de los atributos en muchas de las instancias que las representan, y estos modelos se ven incapaces de extraer suficiente información para diferenciar las clases. Es decir, al utilizar esta representación, únicamente la primera instancia de una serie secuencial muy larga (nótese cómo ‘Dormir’ o ‘Salir’ son actividades de amplia duración) es la que contiene información relativa a los sensores, las instancias del resto de dicha secuencia, para las actividades nombradas, tendrán todos sus atributos (la activación de los sensores) con valor cero. Por tanto, cuando se emplea la representación *ChangePoint*, al existir la misma instancia para diferentes clases, los modelos puramente discriminativos por sí solos, no son capaces de llevar a cabo una generalización adecuada.

Tal y como se puede apreciar en las tablas, los valores para la desviación estándar

Tabla 4.3: Resultados experimentales para el conjunto de datos ‘KasterenA’. Se muestra el valor de *F-Measure* promediado tras la validación cruzada (expresado en %).

| Representación | Conjunto de datos KasterenA | | | | | | | |
|-----------------|-----------------------------|------------------|---------|----------------|---------|---------|---------------|---------|
| | HMM | Modelos Híbridos | | Clasificadores | | | | |
| | | RNA | SVM | RNA | SVM | Árboles | k-NN (k=5) | Reglas |
| Raw | 41 ± 20 | 56 ± 10 | 67 ± 11 | 55 ± 11 | 58 ± 11 | 58 ± 11 | 57 ± 11 | 52 ± 11 |
| ChangePoint | 72 ± 14 | 61 ± 9 | 79 ± 9 | 53 ± 9 | 56 ± 11 | 59 ± 10 | 56 ± 9 | 53 ± 12 |
| LastSensor | 59 ± 15 | 68 ± 10 | 69 ± 8 | 64 ± 9 | 65 ± 8 | 65 ± 8 | 65 ± 9 | 67 ± 9 |
| Promedio | 57 ± 16 | 62 ± 10 | 72 ± 9 | 57 ± 10 | 60 ± 10 | 61 ± 10 | 59 ± 10 | 57 ± 11 |

Tabla 4.4: Resultados experimentales para el conjunto de datos ‘KasterenB’. Se muestra el valor de *F-Measure* promediado tras la validación cruzada (expresado en %).

| Representación | Conjunto de datos KasterenB | | | | | | | |
|-----------------|-----------------------------|------------------|--------|----------------|---------|---------|---------------|---------|
| | HMM | Modelos Híbridos | | Clasificadores | | | | |
| | | RNA | SVM | RNA | SVM | Árboles | k-NN (k=5) | Reglas |
| Raw | 39 ± 13 | 49 ± 11 | 50 ± 9 | 50 ± 11 | 56 ± 11 | 50 ± 13 | 54 ± 12 | 48 ± 13 |
| ChangePoint | 51 ± 16 | 67 ± 8 | 74 ± 8 | 60 ± 9 | 60 ± 6 | 59 ± 8 | 58 ± 8 | 55 ± 8 |
| LastSensor | 40 ± 17 | 66 ± 7 | 57 ± 8 | 63 ± 10 | 65 ± 10 | 65 ± 10 | 62 ± 9 | 60 ± 9 |
| Promedio | 43 ± 15 | 61 ± 9 | 60 ± 8 | 58 ± 10 | 60 ± 9 | 58 ± 11 | 58 ± 10 | 54 ± 10 |

Tabla 4.5: Resultados experimentales para el conjunto de datos ‘KasterenC’. Se muestra el valor de *F-Measure* promediado tras la validación cruzada (expresado en %).

| Representación | Conjunto de datos KasterenC | | | | | | | |
|-----------------|-----------------------------|------------------|---------|----------------|--------|---------|---------------|---------|
| | HMM | Modelos Híbridos | | Clasificadores | | | | |
| | | RNA | SVM | RNA | SVM | Árboles | k-NN (k=5) | Reglas |
| Raw | 15 ± 8 | 47 ± 13 | 45 ± 8 | 48 ± 12 | 50 ± 8 | 47 ± 9 | 48 ± 9 | 42 ± 12 |
| ChangePoint | 45 ± 8 | 60 ± 8 | 60 ± 10 | 49 ± 5 | 50 ± 9 | 49 ± 7 | 50 ± 6 | 48 ± 8 |
| LastSensor | 46 ± 12 | 65 ± 8 | 60 ± 7 | 62 ± 8 | 65 ± 8 | 62 ± 8 | 65 ± 6 | 64 ± 8 |
| Promedio | 35 ± 9 | 57 ± 10 | 55 ± 8 | 53 ± 8 | 55 ± 9 | 53 ± 8 | 54 ± 7 | 51 ± 9 |

Tabla 4.6: Resultados experimentales para el conjunto de datos ‘OrdonezA’. Se muestra el valor de *F-Measure* promediado tras la validación cruzada (expresado en %).

| Representación | Conjunto de datos OrdonezA | | | | | | | |
|-----------------|----------------------------|------------------|--------|----------------|--------|---------|---------------|---------|
| | HMM | Modelos Híbridos | | Clasificadores | | | | |
| | | RNA | SVM | RNA | SVM | Árboles | k-NN (k=5) | Reglas |
| Raw | 51 ± 7 | 80 ± 7 | 81 ± 7 | 79 ± 5 | 81 ± 5 | 81 ± 6 | 78 ± 7 | 79 ± 5 |
| ChangePoint | 56 ± 5 | 65 ± 6 | 62 ± 7 | 58 ± 6 | 57 ± 6 | 58 ± 7 | 54 ± 7 | 49 ± 7 |
| LastSensor | 54 ± 7 | 72 ± 10 | 69 ± 9 | 71 ± 8 | 69 ± 7 | 72 ± 8 | 70 ± 8 | 66 ± 10 |
| Promedio | 54 ± 6 | 72 ± 8 | 71 ± 8 | 70 ± 6 | 69 ± 6 | 71 ± 7 | 67 ± 7 | 65 ± 7 |

Tabla 4.7: Resultados experimentales para el conjunto de datos ‘OrdonezC’. Se muestra el valor de *F-Measure* promediado tras la validación cruzada (expresado en %).

| Representación | Conjunto de datos OrdonezC | | | | | | | |
|-----------------|----------------------------|------------------|--------|----------------|--------|---------|---------------|--------|
| | HMM | Modelos Híbridos | | Clasificadores | | | | |
| | | RNA | SVM | RNA | SVM | Árboles | k-NN (k=5) | Reglas |
| Raw | 69 ± 7 | 74 ± 7 | 76 ± 7 | 69 ± 7 | 72 ± 6 | 70 ± 8 | 69 ± 7 | 69 ± 8 |
| ChangePoint | 65 ± 8 | 62 ± 8 | 68 ± 5 | 55 ± 7 | 55 ± 5 | 56 ± 7 | 53 ± 7 | 50 ± 6 |
| LastSensor | 62 ± 6 | 70 ± 7 | 72 ± 7 | 71 ± 8 | 70 ± 7 | 72 ± 8 | 71 ± 8 | 68 ± 7 |
| Promedio | 65 ± 7 | 69 ± 7 | 72 ± 6 | 65 ± 7 | 66 ± 6 | 66 ± 8 | 64 ± 7 | 62 ± 7 |

dar son bastante elevados. Esto es debido a que existen grandes diferencias en los datos de actividad recogidos para los diferentes días que componen los conjuntos, y consecuentemente existen grandes diferencias en los datos que componen cada una de las iteraciones de la validación cruzada. Por ejemplo, de lunes a viernes los usuarios suelen salir más a menudo, mientras que durante los fines de semana pasan mucho más tiempo dentro de la vivienda.

En lo relativo a los modelos empleados, ninguno de los clasificadores ofrece consistentemente el mejor rendimiento para ninguna de las representaciones usadas. Aunque de un modo general las SVM obtengan los valores más altos y los clasificadores basados en reglas los resultados más pobres, las diferencias mostradas en su gran mayoría no son estadísticamente significativas.

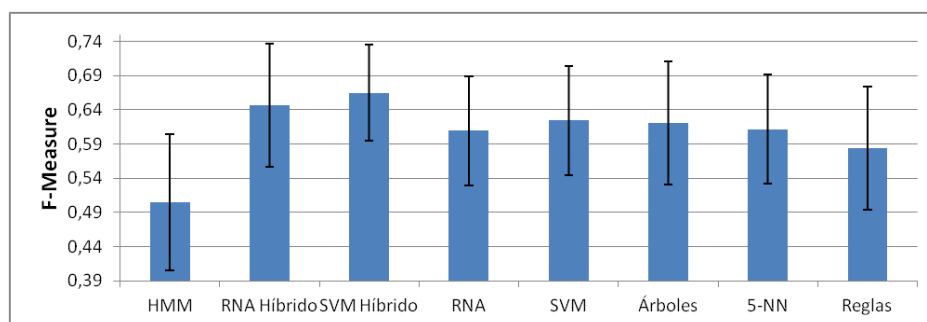


Figura 4.9: Comparativa promediada para todos los modelos empleados en el análisis de métodos híbridos.

Donde sí se aprecia un mejor rendimiento es en los modelos híbridos. Aunque hay casos particulares donde el modelo híbrido no mejora a su equivalente independiente (por ejemplo, las SVM para los conjuntos de datos ‘KasterenB’ y ‘KasterenC’), si se presta atención a los valores promediados que se muestran en la Figura 4.9, se puede apreciar cómo los modelos híbridos mejoran significativamente tanto a su versión puramente discriminativa como a los modelos generativos. Por ejemplo, la diferencia en rendimiento entre el modelo híbrido HMM/SVM y el modelo discriminativo SVM es significativa según los dos test estadísticos empleados. El modelo generativo hibridizado con las redes de neuronas ofrece unos resultados similares, mejorando tanto su versión independiente como el HMM. El hecho de que los modelos híbridos sean capaces de mejorar los resultados que obtienen sus versiones no híbridas es plenamente consistente con la literatura.

Al comparar las dos propuestas híbridas entre ellas las diferencias no son tan relevantes. Aunque el HMM combinado con SVM obtiene unos valores mejores de *F-Measure*, las diferencias son estadísticamente significativas únicamente según el test de significancia con signo de Wilcoxon, pero no según la prueba *t* de Student.

Tabla 4.8: Matriz de confusión del HMM entrenado mediante EM para el conjunto de datos ‘KasterenA’ empleando la representación *LastSensor*.

| Real | Predicho | | | | | | | | |
|-----------------|----------|-----------------|-------|--------|-----------|-------|-------|---------------|-------|
| | Salir | Usar el inodoro | Ducha | Dormir | Desayunar | Cenar | Beber | Sin etiquetar | |
| Salir | 19282 | 12 | 7 | 0 | 0 | 0 | 0 | 30 | 19331 |
| Usar el inodoro | 4 | 259 | 13 | 19 | 0 | 2 | 6 | 46 | 349 |
| Ducha | 1 | 13 | 229 | 0 | 0 | 1 | 0 | 7 | 251 |
| Dormir | 0 | 44 | 0 | 10778 | 0 | 0 | 0 | 3 | 10825 |
| Desayunar | 0 | 3 | 0 | 1 | 31 | 22 | 10 | 11 | 78 |
| Cenar | 0 | 0 | 0 | 0 | 12 | 225 | 10 | 13 | 260 |
| Beber | 0 | 1 | 1 | 0 | 1 | 10 | 29 | 5 | 47 |
| Sin etiquetar | 309 | 517 | 401 | 196 | 61 | 861 | 820 | 915 | 4080 |
| | 19596 | 849 | 651 | 10994 | 105 | 1121 | 875 | 1030 | |

Un resultado especialmente notable es el rendimiento de los HMM, que en comparación con los modelos híbridos ofrecen un resultado significativamente inferior. La diferencia en rendimiento entre modelos generativos y discriminativos principalmente se debe al método empleando para la estimación de sus parámetros. Los parámetros del HMM son calculados para cada actividad separadamente, mientras que los parámetros de los modelos discriminativos son calculados para optimizar la clasificación de la secuencia de datos como un todo. El método de optimización de los modelos puramente discriminativos favorece la clasificación de las clases más frecuentes, en detrimento de las clases menos comunes, por lo que la clasificación de estos modelos está típicamente sesgada hacia las actividades más representadas en los conjuntos.

Para poder ver más claramente las diferencias en el rendimiento de estas propuestas, se muestran las matrices de confusión de tres de estos modelos. La matriz del modelo puramente generativo (HMM) se muestra en la Tabla 4.8, la del modelo discriminativo con mejor rendimiento (SVM) se muestra en la Tabla 4.9 y la del modelo híbrido con mejor rendimiento (HMM/SVM) en la Tabla 4.10. Para poder llevar a cabo una comparativa adecuada, todas estas tablas se han obtenido con el conjunto de datos ‘KasterenA’ empleado la representación *LastSensor*.

Tal y como se puede apreciar en la Tabla 4.9, el modelo discriminativo no clasifica correctamente las actividades menos comunes (‘Beber’ y ‘Desayunar’), a cambio de favorecer la clasificación de aquellas actividades que mayor peso tienen dentro del conjunto de datos (‘Salir’). Como consecuencia, este tipo de modelo obtiene un buen rendimiento si se tiene en consideración un promedio de todas las actividades, es decir, si se considera como métrica el número global de instancias bien clasificadas dentro del conjunto total, pero hay que tener en cuenta que las clases menos comunes se ven desfavorecidas.

Por otro lado, el HMM tiene menos problemas identificando las clases menos

Tabla 4.9: Matriz de confusión del modelo SVM para el conjunto de datos ‘KasterenA’ empleando la representación *LastSensor*.

| Real | Predicho | | | | | | | | |
|-----------------|--------------|-----------------|------------|--------------|-----------|------------|----------|---------------|-------|
| | Salir | Usar el inodoro | Ducha | Dormir | Desayunar | Cenar | Beber | Sin etiquetar | |
| Salir | 19307 | 2 | 3 | 15 | 0 | 0 | 2 | 2 | 19331 |
| Usar el inodoro | 9 | 237 | 7 | 87 | 0 | 0 | 0 | 9 | 349 |
| Ducha | 0 | 4 | 156 | 85 | 0 | 0 | 0 | 6 | 251 |
| Dormir | 92 | 28 | 16 | 10685 | 0 | 0 | 0 | 4 | 10825 |
| Desayunar | 0 | 2 | 1 | 5 | 16 | 8 | 3 | 43 | 78 |
| Cenar | 1 | 4 | 0 | 1 | 4 | 111 | 8 | 131 | 260 |
| Beber | 1 | 1 | 0 | 0 | 3 | 6 | 0 | 36 | 47 |
| Sin etiquetar | 321 | 13 | 29 | 489 | 7 | 12 | 2 | 3207 | 4080 |
| | 19731 | 291 | 212 | 11367 | 30 | 137 | 15 | 3438 | |

Tabla 4.10: Matriz de confusión del modelo híbrido HMM/SVM para el conjunto de datos ‘KasterenA’ empleando la representación *LastSensor*.

| Real | Predicho | | | | | | | | |
|-----------------|--------------|-----------------|------------|--------------|-----------|------------|-----------|---------------|-------|
| | Salir | Usar el inodoro | Ducha | Dormir | Desayunar | Cenar | Beber | Sin etiquetar | |
| Salir | 19294 | 2 | 3 | 28 | 0 | 0 | 2 | 2 | 19331 |
| Usar el inodoro | 91 | 183 | 7 | 59 | 0 | 0 | 0 | 9 | 349 |
| Ducha | 12 | 12 | 201 | 20 | 0 | 0 | 0 | 6 | 251 |
| Dormir | 92 | 18 | 14 | 10697 | 0 | 0 | 0 | 4 | 10825 |
| Desayunar | 0 | 6 | 1 | 11 | 15 | 7 | 3 | 35 | 78 |
| Cenar | 0 | 4 | 0 | 1 | 4 | 186 | 8 | 57 | 260 |
| Beber | 1 | 1 | 0 | 0 | 3 | 8 | 20 | 14 | 47 |
| Sin etiquetar | 121 | 113 | 129 | 289 | 7 | 112 | 2 | 3307 | 4080 |
| | 19611 | 339 | 355 | 11105 | 29 | 313 | 35 | 3434 | |

comunes, pero obtiene un peor rendimiento general. En la Tabla 4.8 se ve cómo este modelo mejora el reconocimiento de las actividades poco frecuentes, pero debido a que no posee la capacidad de generalizar de los modelos discriminativos, en algunas actividades comete muchos fallos. Especialmente reseñable es el caso de la clase ‘Sin etiquetar’, donde la clasificación del modelo prácticamente sigue un esquema aleatorio, debido a las peculiaridades de las instancias que conforman dicha clase, donde el espacio de valores de los atributos no posee un patrón definido. También se aprecia cómo el HMM confunde bastante varias actividades que se llevan a cabo típicamente en la misma habitación (y por tanto con sensores relacionados), como es en este caso las actividades ‘Beber’, ‘Cenar’ y ‘Desayunar’. Este ejemplo de actividades modeladas mediante sensores relacionados pone de manifiesto las deficiencias de la propuesta generativa, en la que se asume la independencia de los atributos.

Sin embargo, el modelo híbrido HMM/SVM consigue solventar notablemente las deficiencias que presentan las dos propuestas anteriores. Es bastante ilustrativa la Tabla 4.10, en donde se aprecia cómo este modelo obtiene una clasificación similar a la SVM en lo que respecta a las clases más comunes, pero teniendo en mucha más consideración las actividades menos frecuentes. El añadir la información temporal de los HMM (la probabilidad de transición entre actividades) a la clasificación del modelo discriminativo provoca principalmente un mejor rendimiento del modelo en lo que respecta a las actividades poco presentes.

4.4. Inferencia Bayesiana en HMM

En este apartado se presenta un método de aprendizaje para los HMM basado en simulación estocástica, como alternativa a los algoritmos de inferencia clásicos empleados en anteriores propuestas del dominio [132]. El objetivo de esta propuesta es evaluar si aplicando un enfoque bayesiano bajo un esquema de simulación estocástica para estimar los parámetros de un HMM se pueden evitar los problemas que típicamente acompañan a los métodos clásicos de inferencia en el dominio, a saber: rendimiento condicionado a los valores iniciales y tendencia a converger hacia máximos locales.

En un enfoque probabilístico clásico, el objetivo de la inferencia estadística consiste en definir una hipótesis sobre un conjunto de datos y , en relación al modelo probabilístico $p(y|\theta_0)$ que podría haberlos generado, en donde θ_0 son hipotéticamente los valores ciertos para el parámetro θ .

En contraste con este enfoque, la inferencia bayesiana introduce como parte del modelo una distribución a priori $p(\theta)$. En esencia, la inferencia bayesiana está basada en la distribución de probabilidad del parámetro dados los datos, en lugar de la distribución de los datos dado el parámetro, y esta diferencia conduce a inferencias mucho más naturales. Lo único que se requiere para el proceso de inferencia bayesiana es la especificación previa de una distribución a priori de probabilidad, la cual representa el conocimiento acerca del parámetro antes de obtener cualquier información respecto a los datos. Esto es, dicha distribución representa un estado de conocimiento o ignorancia acerca de θ antes de que los datos sean obtenidos. Una vez se dispone de la distribución a priori, el modelo probabilístico $p(y|\theta)$ y el conjunto de datos y , es posible calcular la distribución de probabilidad $p(\theta|y)$, es decir, la distribución a posteriori.

La interpretación bayesiana para nuestra tarea de clasificación de una secuencia de observaciones, comienza considerando todas las posibles hipótesis (secuencias de estados) para poder seleccionar aquella más probable dada la secuencia de observaciones $x_{1:T}$. Esto es, se desea encontrar, de entre todas las posibles secuencias de variables ocultas $y_{1:T}$, aquella que maximice la probabilidad $p(y_{1:T}|x_{1:T})$.

Es por esto que a la hipótesis más probable se le denomina hipótesis *maximum a posteriori* o MAP.

En lo relativo a los HMM, donde el problema de la clasificación de una secuencia es la búsqueda del modelo que mejor clasifique dicha secuencia, la fórmula del *maximum a posteriori* se define como:

$$\hat{\theta}_{MAP} = \underset{\theta_i}{\operatorname{argmax}} p(\theta | \mathbf{x}_{1:T}) \quad (4.19)$$

donde la notación $\hat{\cdot}$ significa que la secuencia MAP es una estimación y θ representa el conjunto de parámetros ($\theta = \{\pi, \mathbf{A}, \mathbf{B}\}$) que definen el modelo dinámico. El razonamiento probabilístico tiene lugar a través del cálculo de la probabilidad condicional del modelo dada una secuencia de observaciones $\mathbf{x}_{1:T}$. Pero, a pesar de que la fórmula del *maximum a posteriori* garantiza una secuencia óptima, no se sabe cómo calcular $p(\theta | \mathbf{x}_{1:T})$. La regla de Bayes (ver Función 4.1) permite transformar esa ecuación en un conjunto de probabilidades que resultan más sencillas de calcular, y el problema de la clasificación se puede resolver maximizando la siguiente fórmula:

$$\hat{\theta}_{MAP} = \underset{\theta_i}{\operatorname{argmax}} p(\mathbf{x}_{1:T} | \theta) p(\theta) \quad (4.20)$$

El primer término corresponde a la verosimilitud de la secuencia de observaciones dado el modelo, y el segundo a la probabilidad a priori del modelo. La secuencia de variables ocultas será aquella que maximice $p(\mathbf{y}_{1:T} | \mathbf{x}_{1:T}, \hat{\theta})$.

Sin embargo, existen ocasiones en donde no es posible encontrar las distribuciones a posteriori de los parámetros en su forma analítica cerrada. Para situaciones así, donde no aplica el enfoque bayesiano clásico, son muy populares los esquemas de simulación estocástica, como por ejemplo el método Montecarlo basado en cadenas de Markov (MCMC, por sus siglas en inglés) [150]. Estos métodos se han demostrado útiles para realizar el análisis a posteriori de los HMM, permitiendo identificar la distribución conjunta de las variables ocultas directamente. Dicho análisis ofrece un método muy robusto frente a valores atípicos, y se define mediante la ecuación:

$$p(\mathbf{y}_{1:T} | \mathbf{x}_{1:T}) = \int p(\mathbf{y}_{1:T} | \mathbf{x}_{1:T}, \theta) p(\theta | \mathbf{x}_{1:T}) d\theta \quad (4.21)$$

Para ver porqué es útil integrar sobre todos los posibles parámetros usando estadística bayesiana se puede emplear el conocido ejemplo de la moneda trucada. Supongamos que existe una moneda que puede estar trucada ($y = 1$) o no ($y = 0$), con una probabilidad del 50 %. En este caso θ es la probabilidad de obtener una cara (cara se representa mediante el símbolo C y cruz mediante X). Si la moneda está trucada, asumimos una distribución uniforme sobre θ , sino $\theta = 0,5$.

Al obtener la observación \mathbf{x} , que es el resultado de tirar la moneda 10 veces, queremos estimar si la moneda está trucada o no, es decir, queremos determinar y . Suponiendo también que existe una distribución a priori uniforme sobre θ , donde $p(\theta) = 1$ para todo $\theta \in [0, 1]$, primero se aplica una metodología estándar para llevar a cabo un aprendizaje MAP de θ , con el que determinar un valor de \mathbf{x} que maximice $p(\mathbf{x}|\mathbf{y}, \hat{\theta})$. Para este ejemplo, un cálculo elemental muestra que dicho aprendizaje MAP dará como resultado $\hat{\theta} = n_C/10$, donde n_C es el número de caras que hemos obtenido en nuestras observaciones (y por tanto n_X representará el número de cruces). Como consecuencia $p(\mathbf{y}|\mathbf{x}, \hat{\theta})$ tenderá a $y = 1$ cuando la secuencia no contenga exactamente 5 caras, y por lo tanto, se estimará que la moneda está trucada. Del mismo modo, esta estimación optará por $y = 0$ cuando el número de caras en la secuencia sea exactamente 5, ofreciendo un resultado que realmente no representa la realidad.

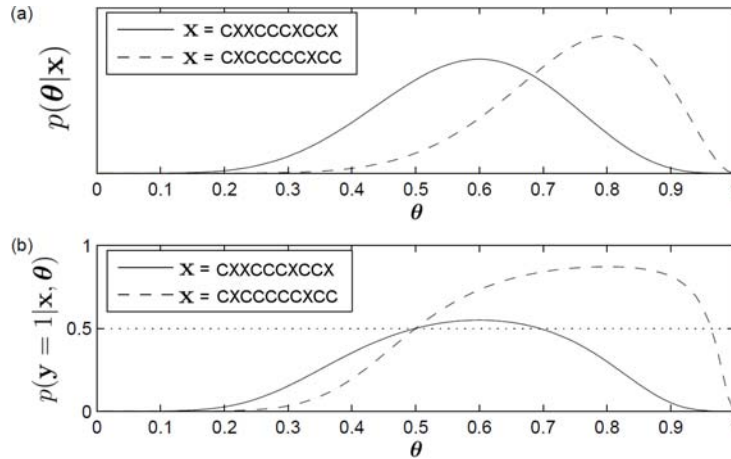


Figura 4.10: Propuesta bayesiana para estimar la variable oculta y , a partir de una variable observable \mathbf{x} . (a) Distribución a posteriori del parámetro θ dado \mathbf{x} . Probabilidad para $y = 1$ dado las observaciones \mathbf{x} y el parámetro θ .

Por otro lado, si empleamos la distribución de las variables ocultas que ofrece el análisis bayesiano, tenemos que la probabilidad para una moneda trucada es:

$$p(y = 1|\mathbf{x}) = 1 / \left(1 + \frac{11!}{n_C!n_X!2^{10}} \right) \quad (4.22)$$

Esta probabilidad es significativamente inferior a 0,5 cuando $n_C = 5$, y sólo tiende a $y = 1$ (moneda trucada) en las secuencias donde $n_X \geq 8$ o donde $n_X \leq 2$. Esta aproximación ofrece unos resultados más razonables, permitiendo un cálculo más robusto del valor de y en función de la estimación de θ . En la Figura 4.10(a) se puede apreciar cómo para una secuencia de observaciones donde $n_C = 6$, una estimación MAP de los parámetros obtiene un valor $\hat{\theta} = 0,6$, y la probabilidad

de que la moneda esté trucada es mayor que 0,5 sólo para un pequeño rango de θ alrededor de $\hat{\theta}$ (Figura 4.10(b)). Por el contrario, una secuencia de observaciones donde $n_C = 8$ tiende a $y = 1$ (estima que la moneda estará trucada) para un amplio rango de valores de θ alrededor de $\hat{\theta}$.

En la literatura existen diferentes estudios que se decantan por este tipo de procesos de inferencia bayesianos para los HMM (también conocidos como *fully bayesian*), en detrimento de otras propuestas más clásicas como puede ser el algoritmo de máxima expectación. Esto es debido a que combinan de forma eficiente la estimación de los parámetros y de la topología del modelo [209].

Son comunes las propuestas para aplicar este tipo de métodos a los problemas de reconocimiento del habla [110]. Goldwater *et. al.* propusieron un HMM para realizar etiquetado gramatical entrenado mediante un método de inferencia totalmente bayesiano, con el que obtuvieron una mejora muy significativa frente a los métodos de máxima verosimilitud [87]. Zhang *et. al.* recientemente han presentado un estudio similar, en el que proponen un HMM adaptativo para reconocer caracteres chinos. En su trabajo demuestran cómo el aprendizaje bayesiano se puede emplear, no solamente para estimar los parámetros del modelo, sino también para definir parte de su topología [265].

Este tipo de inferencia *full bayesian* sobre modelos de Markov ha ofrecido buenos resultados en áreas muy diversas. Por ejemplo, en el ámbito de la depuración de código, Pievatolo *et. al.* demostraron cómo el definir la posibilidad de un fallo de programación mediante probabilidades a priori, da como resultado un sistema flexible que es capaz de modelar los intervalos de tiempo entre errores adecuadamente [188].

Nur *et. al.* realizaron un estudio acerca de cuales deben ser la características que gobiernen las probabilidades a priori para estos modelos de aprendizaje. En concreto, demostraron la importancia de elegir inicialmente unos valores no informativos para los parámetros de las distribuciones a priori, enmarcándolo dentro del problema de reconocimiento de secuencias de ADN [174].

También se ha demostrado la utilidad de estas propuestas en el campo de la genómica [262, 89]. En general, se han demostrado útiles en este área los algoritmos para generar muestras aleatorias, y en particular, el algoritmo de muestreo de Gibbs es el que según los autores ha ofrecido mayor flexibilidad y rendimiento.

Un elemento común en la mayoría de estos trabajos, es el hecho de que este enfoque bayesiano suele destacar por encima de otros tipos de inferencia más clásicos, como pueden ser los algoritmos de esperanza-maximización. Los métodos de esperanza-maximización o EM presentan varios problemas bien conocidos: son algoritmos numéricamente sensibles, dependen fuertemente de las estimaciones iniciales de los parámetros del modelo y su tipo de búsqueda suele conducir a soluciones subóptimas del problema [208, 96].

Incluso existen estudios cuya finalidad es comparar los métodos de esperanza-

maximización típicamente empleados en la inferencia de los HMM (el algoritmo *Baum-Welch*) con este tipo de inferencia totalmente bayesiana [211]. Dicho estudio, presentado por Ryden *et. al.*, concluye que los modelos bayesianos en combinación con los esquemas de simulación estocástica presentan ciertas ventajas cuando se ha de tratar con tipos de modelos o problemas inferenciales complejos. Sin embargo, cuando se deben modelar problemas más simples ambas propuestas presentan inconvenientes, ya que los algoritmos EM convergen a máximos locales y el muestreo de las técnicas basadas en MCMC es muy pobre. No obstante, los autores demostraron cómo los algoritmos de inferencia bayesiana mediante simulación estocástica, requerían de un menor tiempo de computación.

Otros trabajos en esa misma línea han demostrado cómo las técnicas de simulación estocástica, gracias a su flexibilidad y carácter no determinista, tienen menos problemas de convergencia cuando el proceso de inferencia comienza con la misma situación paramétrica [67]. En lo relativo a su complejidad, los estimadores bayesianos mediante MCMC consumen un tiempo $O(n)$ en cada iteración, mientras que los algoritmos de máxima expectación (y en concreto el algoritmo *Baum-Welch*) requiere de un tiempo $O(n^2)$ para cada una de sus iteraciones [122].

En esta tesis doctoral, postulamos que este tipo de inferencia bayesiana es más adecuado para la estimación paramétrica de los HMM en relación a nuestro dominio. La flexibilidad que ofrecen las técnicas de simulación estocástica permite la obtención de modelos generativos más precisos que aquellos que se pueden obtener mediante técnicas más clásicas, como son los algoritmos de máxima expectación. Hasta donde nosotros conocemos, este tipo de aprendizaje sobre el HMM no ha sido aplicado a nuestro dominio.

4.4.1. Simulación Estocástica vía Métodos MCMC

Una de las limitaciones que supone emplear un enfoque bayesiano es que el cálculo de la distribución a posteriori a menudo requiere de la integración de funciones multidimensionales. Esto puede llegar a ser computacionalmente demasiado costoso, pero existen propuestas que permiten la integración sobre distribución conjuntas [229, 78]. Tal y como se ha mencionado anteriormente, una de las más populares son los esquemas de simulación estocástica mediante MCMC, y es el tipo de algoritmo que nosotros aplicamos en esta tesis doctoral.

Los algoritmos MCMC son métodos de simulación para generar muestras de las distribuciones a posteriori y no son exclusivos de la inferencia Bayesiana, sino que pueden usarse para simular valores de una distribución de la que no es fácil generar muestras [234]. Son dos las razones principales por las cuales el método Montecarlo basado en cadenas de Markov es una opción adecuada a nuestro contexto:

- Permite una gran flexibilidad en el modelado. Realmente, en nuestro dominio, para obtener la secuencia de actividades más probable se necesitaría integrar sobre todas las posibles secuencias que pudieran existir. Llevar a cabo dicho proceso de forma analítica puede ser numéricamente complicado o directamente intratable. Sin embargo, los MCMC permiten estimar integrales en espacios dimensionales muy grandes mediante la técnica de muestreo.
- Cuando se aplican bajo un enfoque bayesiano permite el análisis de todos los parámetros del modelo. Permiten incluir la cadena oculta dentro del proceso de aprendizaje.

Distribuciones a priori

La distribución a priori cumple un papel importante en el análisis bayesiano ya que mide el grado de conocimiento inicial que se tiene de los parámetros en estudio. Pueden verse como el mecanismo que tenemos para incorporar información a los parámetros del modelo, y en nuestro caso, representan el conocimiento que tenemos previamente del comportamiento del usuario. Si bien su influencia disminuye a medida que más información muestral está disponible, el uso de una u otra distribución a priori determinará ciertas diferencias en la distribución a posteriori. Sin embargo, tal y como se ha demostrado en algunos estudios [174], cuando dicho conocimiento es nulo o limitado la selección de una distribución a priori adecuada adquiere una connotación especial, pues será necesario elegir una distribución a priori que no influya sobre ninguno de los posibles valores de los parámetros en cuestión. Estas distribuciones a priori reciben el nombre de difusas o no informativas.

Como se puede suponer, la elección de dichas distribuciones a priori no es un punto trivial, pues sirven para caracterizar una situación de la que se conoce poco (en algunos casos, nada). Un criterio muy común es el de emplear las distribuciones conjugadas. Por familia conjugada de distribuciones a priori se debe entender una familia de dichas distribuciones que, cuando son combinadas con la función de verosimilitud mediante el teorema de Bayes, resulta una distribución a posteriori que es de la misma familia paramétrica de la distribución a priori [19, 22]. Esto es, la distribución a priori y la a posteriori son la misma (cambiando los parámetros al incorporar la verosimilitud de los datos). Para nuestro sistema bayesiano las distribuciones a priori son seleccionadas empleando este concepto de distribuciones conjugadas.

Los parámetros que gobiernan las distribuciones a priori son comunmente denominados hiperparámetros, y en esta tesis doctoral al conjunto de hiperparámetros lo denotamos mediante Ψ , para distinguirlo claramente del conjunto de pa-

| Factores | Probabilidades | | Distribuciones a priori | |
|----------------|----------------|--------------|-------------------------|-----------------|
| | Nombre | Parámetros | Nombre | Hiperparámetros |
| Estado inicial | Multinomial | π | Dirichlet | η |
| Transición | Multinomial | \mathbf{A} | Dirichlet | ρ |
| Observación | Binomial | \mathbf{B} | Beta | ω, ν |

Tabla 4.11: Resumen de las probabilidades empleadas en el HMM, definidas por el conjunto de parámetros $\theta = \{\pi, \mathbf{A}, \mathbf{B}\}$, y sus correspondientes distribuciones a priori, definidas por los hiperparámetros $\Psi = \{\eta, \rho, \omega, \nu\}$.

rámetros θ . Un resumen de las diferentes distribuciones y sus correspondientes parámetros puede apreciarse en la Tabla 4.11.

Implementación mediante muestreo de Gibbs

El muestreo de Gibbs es uno de los métodos más empleados dentro de los esquemas MCMC para generar variables aleatorias indirectamente de una distribución sin tener que calcular la densidad. En esta tesis doctoral, adoptamos este enfoque para llevar a cabo una estimación eficiente del modelo bayesiano.

Primero supongamos un conjunto de parámetros $\theta = (\theta_1, \dots, \theta_k)$, donde para todo $i = 1, \dots, k$, es posible simular su distribución a posteriori. Podemos representar un conjunto de distribuciones univariadas condicionales para cada uno de los parámetros que componen θ , tal que:

$$\begin{aligned}
 & p(\theta_1 | \theta_2, \theta_3, \dots, \theta_k, \mathbf{y}) \\
 & p(\theta_2 | \theta_1, \theta_3, \dots, \theta_k, \mathbf{y}) \\
 & \vdots \\
 & p(\theta_k | \theta_1, \theta_2, \dots, \theta_{k-1}, \mathbf{y})
 \end{aligned} \tag{4.23}$$

Mediante el empleo del muestreo de Gibbs, las distribuciones condicionales son usadas para generar secuencias de valores de los parámetros aleatorios univariados, para cada uno de los elementos en θ . Dadas esas distribuciones posteriores condicionales, el muestreo de Gibbs se inicia fijando un valor inicial a los parámetros θ^0 , con el cual comienza el proceso.

A partir de este momento, el siguiente conjunto de valores paramétricos son secuencialmente generados para muestrear los elementos de θ a partir de las distribuciones condicionales:

$$\begin{aligned}
& p(\theta_1^1 | \theta_2^0, \theta_3^0, \dots, \theta_k^0, \mathbf{y}) \\
& p(\theta_2^1 | \theta_1^1, \theta_3^0, \dots, \theta_k^0, \mathbf{y}) \\
& \vdots \\
& p(\theta_k^1 | \theta_1^1, \theta_2^1, \dots, \theta_{k-1}^1, \mathbf{y})
\end{aligned} \tag{4.24}$$

Mediante este mecanismo es posible manejar la complejidad del modelo. En sistemas como el nuestro, donde se emplean HMM, una forma muy común de obtener muestras de las distribuciones a posteriori es incluir las variables ocultas $\mathbf{y}_{1:T}$ en el espacio de estados del modelo. La implementación del muestreo de Gibbs empleada en este trabajo está basada en la presentada en [208], y cada iteración del algoritmo se compone de los siguientes pasos:

- actualización de la matriz \mathbf{A} , que modela la distribución de transición,
- actualización del vector π , que modela la distribución sobre los estados iniciales,
- actualización de la matriz \mathbf{B} , que modela la distribución de las observaciones,
- actualización de las variables ocultas \mathbf{y} ,
- actualización del hiperparámetro ρ ; parámetro de la distribución a priori (Dirichlet) sobre los parámetros de la probabilidad condicional de transición,
- actualización del hiperparámetro η ; parámetro de la distribución a priori (Dirichlet) sobre los parámetros de la probabilidad incondicional de estado,
- actualización de los hiperparámetros ω y ν ; parámetro de la distribución a priori (Beta) sobre los parámetros de la probabilidad condicional de las observaciones,

Siguiendo el esquema propuesto en [208], se obtiene una muestra de cada fila \mathbf{a}_i de la matriz de transición \mathbf{A} desde una distribución Dirichlet $Dir(\rho_1 + n_{i1}, \dots, \rho_k + n_{ik})$, donde:

$$n_{ij} = \sum_{t=2}^T \mathbb{I}_{\{y_{t-1}=i, y_t=j\}} \tag{4.25}$$

es el número de transiciones desde el estado i hacia el estado j ; $\mathbb{I}_{\{\cdot\}}$ denota a la función indicatriz. La muestra del vector que parametriza la distribución sobre los estados iniciales π se obtiene un modo equivalente.

A continuación, se obtiene una muestra de la k -ésima fila de la matriz \mathbf{B} que caracteriza las probabilidades de emisión desde una distribución Beta $Be(\omega_1^k + s_{k1}, \dots, \omega_N^k + s_{kN}, \nu_1^k + r_{k1}, \dots, \nu_N^k + r_{kN})$ donde:

$$s_{kn} = \sum_{t=1}^T \mathbb{I}_{\{y_t=k, x_t^n=1\}} \quad (4.26)$$

$$r_{kn} = \sum_{t=1}^T \mathbb{I}_{\{y_t=k, x_t^n=0\}} \quad (4.27)$$

es respectivamente el número efectivo de observaciones de ceros y unos para el sensor y el estado k .

La solución adoptada para actualizar las variables ocultas es mediante el muestreo de \mathbf{y} directamente desde $p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$, usando una versión estocástica del algoritmo recursivo *forward-backward* [214]. El paso *forward* o “hacia adelante” de este método genera matrices P_2, \dots, P_n , donde $P_t = (p_{tij})$ y $p_{pij} = p(y_{t-1} = i, y_t = j | x_1, \dots, x_t, \boldsymbol{\theta})$. Esto es, P_t es la distribución conjunta de (y_{t-1}, y_t) dado los parámetros del modelo y la secuencia de observaciones hasta el instante t . P_t es calculado empleando P_{t-1} como sigue:

$$\begin{aligned} p_{tij} &\propto p(y_{t-1} = i, y_t = j, x_t | x_1, \dots, x_{t-1}, \boldsymbol{\theta}) \\ &= p(y_{t-1} = i | x_1, \dots, x_{t-1}, \boldsymbol{\theta}) a_{ij} b_j(x_t) \end{aligned} \quad (4.28)$$

manteniendo la proporcionalidad ajustada para $\sum_i \sum_j p_{tij} = 1$. Hay que tener en cuenta que $p(y_t = i | x_1, \dots, x_t, \boldsymbol{\theta}) = \sum_j p_{tij}$ puede ser calculada una vez se conozca P_t , dando lugar por tanto al próximo paso en el algoritmo recursivo.

El primer paso del proceso recursivo mediante el que se define la etapa “hacia atrás” o *backward* es simular el estado oculto y_T empleando $p(y_T | \mathbf{x}, \boldsymbol{\theta})$. Por tanto, y_t es obtenido recursivamente desde la distribución proporcional a la columna y_{t+1} de P_{t+1} . De este modo, la recursión estocástica hacia atrás obtiene muestras usando $p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$ como:

$$p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = p(y_T | \mathbf{x}, \boldsymbol{\theta}) \prod_{t=1}^{T-1} p(y_{T-t} | y_T, \dots, y_{T-t+1}, \mathbf{x}, \boldsymbol{\theta}) \quad (4.29)$$

donde

$$p(y_{T-t} = i | y_T, \dots, y_{T-t+1}, \mathbf{x}, \boldsymbol{\theta}) = p(y_{T-t} = i | y_{T-t+1}, x_1, \dots, x_{T-t+1}, \boldsymbol{\theta}) \\ \propto p_{T-t+1, i, s_{T-t+1}} \quad (4.30)$$

Por lo tanto, y a modo de resumen, en la implementación que usamos del muestreo de Gibbs, se alternan las actualizaciones de los hiperparámetros condicionadas sobre las observaciones y la cadena de variables ocultas, las actualizaciones de los parámetros condicionadas sobre los hiperparámetros y las actualizaciones de la cadena de variables ocultas condicionadas sobre las observaciones y los parámetros.

En general, no es fácil determinar en qué momento los valores simulados han convergido aproximadamente a la distribucon estacionaria. Existen numerosos procedimientos en la literatura para estudiar la convergencia del algoritmo [248]. Un método empírico comúnmente utilizado, y el que hemos empleado en nuestra propuesta, es ejecutar el algoritmo varias veces comenzando en distintos valores iniciales y comprobar si el comportamiento de la distribucion estacionaria es la misma.

4.4.2. Evaluación

La motivación principal de presentar esta propuesta es comparar su rendimiento con los sistemas generativos ya estudiados en la literatura. Por ello, este apartado está enfocado principalmente a comparar la inferencia bayesiana anteriormente detallada con la optimización clásica a través de algoritmos de esperanza-maximización. En este apartado se muestran los resultados de aplicar la inferencia MCMC a modelos HMM, evaluados sobre nuestros conjuntos de datos.

Configuración de la Experimentación

Con el fin de comparar los modelos entrenados mediante inferencia bayesiana con las otras propuestas orientadas al reconocimiento de actividades que presentamos en esta tesis doctoral, la configuración de la experimentación para esta propuesta es totalmente similar a la definida en el Apartado 4.3.2.

Esto es, se han empleado los mismos tipos de representación para los sensores, los mismos conjuntos de datos y la parametrización de la experimentación es la misma (la segmentación del flujo de datos también se realiza con una granularidad de tiempo $\Delta t = 60$ segundos).

Comparativa de los modelos

El rendimiento en términos de *F-Measure* para el HMM, obtenido mediante el algoritmo MCMC, se muestra en la Tabla 4.12. Se puede apreciar cómo los resultados según el modo de representación son variados, no ofreciendo un mejor rendimiento con ningún tipo en concreto, aunque se mantiene la proporcionalidad con los valores obtenidos mediante el HMM entrenado con EM.

Las diferencias entre los valores promedio obtenidos por el HMM entrenado mediante EM y el HMM entrenado mediante MCMC son estadísticamente significativas para todos los conjuntos de datos, según los dos tipos de pruebas estadísticas aplicadas. Es reseñable cómo el HMM entrenado con MCMC supera notablemente al HMM entrenado mediante EM, aunque si se observa su matriz de confusión (Tabla 4.13) se puede apreciar como adolece del mismo tipo de problema.

Tabla 4.12: Comparativa entre el HMM entrenado mediante EM y el HMM entrenado mediante MCMC. Se muestra el valor de *F-Measure* promediado tras la validación cruzada (expresado en %).

| | | Conjuntos de datos | | | | |
|------|-----------------|--------------------|----------------|---------------|---------------|---------------|
| | Representación | KasterenA | KasterenB | KasterenC | OrdonezA | OrdonezC |
| MCMC | Raw | 62 ± 12 | 53 ± 11 | 60 ± 8 | 78 ± 7 | 74 ± 7 |
| | ChangePoint | 75 ± 13 | 78 ± 12 | 62 ± 8 | 69 ± 6 | 72 ± 8 |
| | LastSensor | 61 ± 10 | 64 ± 12 | 67 ± 10 | 69 ± 7 | 69 ± 7 |
| | Promedio | 66 ± 11 | 65 ± 11 | 63 ± 8 | 72 ± 6 | 72 ± 7 |
| EM | Raw | 41 ± 20 | 39 ± 13 | 15 ± 8 | 51 ± 7 | 69 ± 7 |
| | ChangePoint | 72 ± 14 | 51 ± 16 | 45 ± 8 | 57 ± 5 | 65 ± 8 |
| | LastSensor | 59 ± 15 | 40 ± 17 | 46 ± 12 | 54 ± 7 | 62 ± 6 |
| | Promedio | 57 ± 16 | 43 ± 15 | 35 ± 9 | 54 ± 6 | 65 ± 7 |

Observando la matriz de confusión del modelo, detallada en la Tabla 4.13, se puede comprobar cómo este tipo de inferencia permite crear modelos más precisos, pues el aumento de verdaderos positivos aumenta de forma muy relevante. Pero, al igual que sucede con el HMM entrenado mediante EM, la mayor parte de sus fallos ocurren cuando se trata de clasificar clases con mucha presencia en el conjunto, como son ‘Dormir’ o ‘Sin etiquetar’. Aunque dicho inconveniente, se solventa en gran medida gracias al proceso de optimización que permite el método MCMC, que al no ser un mecanismo tan rígido como la EM, permite una búsqueda más flexible de los parámetros.

Tabla 4.13: Matriz de confusión del HMM entrenado mediante MCMC para el conjunto de datos ‘KasterenA’ empleando la representación *LastSensor*.

| Real | Predicho | | | | | | | | |
|-----------------|----------|-----------------|-------|--------|-----------|-------|-------|---------------|-------|
| | Salir | Usar el inodoro | Ducha | Dormir | Desayunar | Cenar | Beber | Sin etiquetar | |
| Salir | 19309 | 2 | 15 | 0 | 0 | 0 | 1 | 4 | 19331 |
| Usar el inodoro | 4 | 308 | 4 | 25 | 0 | 0 | 2 | 16 | 349 |
| Ducha | 0 | 13 | 228 | 8 | 1 | 0 | 0 | 1 | 251 |
| Dormir | 132 | 523 | 853 | 9313 | 0 | 0 | 0 | 4 | 10825 |
| Desayunar | 0 | 0 | 0 | 7 | 40 | 14 | 16 | 1 | 78 |
| Cenar | 1 | 1 | 0 | 0 | 42 | 195 | 17 | 4 | 260 |
| Beber | 1 | 0 | 0 | 1 | 3 | 1 | 39 | 2 | 47 |
| Sin etiquetar | 122 | 394 | 130 | 210 | 244 | 304 | 309 | 2389 | 4080 |
| | 19569 | 1241 | 1230 | 9564 | 330 | 514 | 384 | 2389 | |

4.5. Conclusiones

En este capítulo de la tesis doctoral hemos presentado y evaluado tres propuestas para llevar a cabo el reconocimiento de actividades humanas en entornos monitorizados mediante redes de sensores inalámbricos: dos esquemas híbridos y un modelo generativo entrenado mediante inferencia totalmente bayesiana. Los resultados experimentales han demostrado cómo los modelos HMM híbridos pueden reconocer eficazmente las actividades del usuario, mejorando tanto a los modelos generativos como a los discriminativos, que típicamente se han empleado para este problema.

El flujo de datos es segmentado en intervalos de tiempo regulares, con una granularidad Δt , con el fin de convertir la tarea en un problema de aprendizaje supervisado secuencial. Los valores de los datos son transformados según tres tipos diferentes de representación, propuestos en la literatura. Los modelos son evaluados en términos de *F-Measure*, una métrica que valora la correcta clasificación de todas las clases por igual.

La creación de los modelos HMM híbridos se basa en aplicar una función sobre la salida de los clasificadores que permita transformar su estimación a posteriori de la actividad en una estimación a priori que pueda ser empleada como la probabilidad de observación del HMM. Dichos modelos híbridos han sido evaluados comparándolos con un conjunto de algoritmos puramente discriminativos y con un esquema generativo que ha sido ampliamente utilizado en este dominio, un modelo de Markov entrenado mediante el algoritmo de esperanza-maximización.

Los modelos puramente discriminativos hacen uso de una ventana deslizante para transformar la tarea de clasificación. Hemos comprobado cómo la longitud de dicha ventana no afecta de forma notable al rendimiento de estos modelos y, en nuestro caso particular, la diferencia en rendimiento para ventanas mayores de

tres instancias deja de ser significativa.

Dependiendo del tipo de representación empleado, los modelos discriminativos no son capaces de generalizar adecuadamente, al no disponer de información adicional sobre la secuencialidad de las actividades, e incluso para algunas clases en particular la clasificación mantiene un esquema aleatorio.

La principal diferencia entre los modelos generativos y los discriminativos radica en el método de aprendizaje. Al clasificar la secuencia de instancias como un todo, los modelos discriminativos tienden a favorecer a las clases más comunes del conjunto, lo que se traduce en que obtiene una tasa de reconocimiento muy pobre sobre las actividades menos frecuentes en el día a día de una persona. Por contra, los modelos discriminativos no penalizan a las actividades menos presentes, pero al modelar cada sensor de forma independiente, confunden muy frecuentemente las actividades que se modelan mediante sensores relacionados, como aquellas que se llevan a cabo en una misma estancia de la vivienda ('Cenar', 'Desayunar' y 'Beber', por ejemplo).

Los modelos híbridos presentados consiguen solventar dichos problemas, combinando la capacidad de generalizar de los modelos discriminativos con la información temporal obtenida de los HMM. Ambos modelos híbridos obtienen mejores resultados que los clasificadores que los componen, esto es, el modelo HMM/RNA consigue mejorar a la RNA, y del mismo modo, el modelo híbrido HMM/SVM obtiene mejores resultados que el clasificador SVM, demostrando cómo añadir información temporal a la clasificación mejora el rendimiento del proceso.

También se ha demostrado cómo el rendimiento del modelo generativo puede mejorar muy notablemente si se emplea inferencia bayesiana para estimar los parámetros. Implementando una versión del algoritmo MCMC se puede llevar a cabo un aprendizaje que ofrece una búsqueda mucho más completa en el espacio de soluciones, y permite integrar la estimación de la cadena oculta dentro del proceso de inferencia. Como resultado se obtiene un clasificador que ofrece un rendimiento significativamente mejor que su homólogo entrenado mediante el algoritmo de esperanza-maximización. El algoritmo EM resulta muy poco eficiente para el aprendizaje de los parámetros del HMM en nuestro dominio, debido a la longitud de nuestras secuencias y a la dimensionalidad de los datos. Aunque el modelo entrenado mediante MCMC conserva las deficiencias propias de los sistemas generativos, obtiene de promedio un rendimiento mucho mejor.

En la comparativa global el modelo híbrido HMM/SVM y el HMM entrenado mediante MCMC obtienen los mejores resultados. En la Tabla 4.14 se aprecia cómo las diferencias entre ambas propuestas no son estadísticamente significativas, de acuerdo al resultado de las dos pruebas de significancia estadística empleadas, con un nivel de significación de 0.05. Sin embargo, cada propuesta presenta un esquema de aprendizaje diferente, y esto debe tenerse en cuenta de cara a su

integración dentro de un sistema de monitorización real. El empleo del modelo HMM entrenado mediante MCMC puede ocasionar que actividades con un espacio de valores parecido se confundan, dando como resultado, por ejemplo, que el sistema no diferencie bien entre la actividad ‘Cenar’ y la actividad ‘Desayunar’, dado que ambas se relacionan con sensores similares. Por su parte, el modelo híbrido tiende a penalizar las actividades poco frecuentes, y hay que valorar si este factor puede ocasionar problemas relevantes, ya que en la práctica una actividad poco frecuente para una persona -comparativamente al resto de actividades del conjunto- podría ser tomarse la medicación.

Tabla 4.14: Valores promediados de las diferentes métricas para cada uno de los modelos propuestos (expresados en %).

| Conjuntos de datos | HMM | | | | | | HMM Híbridos | | | | | |
|--------------------|---------------|-----|-----------|-----------------|-----|-----------|--------------|-----|-----------|-----|-----|-----------|
| | Inferencia EM | | | Inferencia MCMC | | | MLP | | | SVM | | |
| | PPV | VPR | F-Measure | PPV | VPR | F-Measure | PPV | VPR | F-Measure | PPV | VPR | F-Measure |
| KasterenA | 54 | 63 | 57 | 70 | 63 | 66 | 67 | 58 | 62 | 74 | 70 | 72 |
| KasterenB | 39 | 50 | 43 | 68 | 63 | 65 | 65 | 58 | 61 | 61 | 60 | 60 |
| KasterenC | 32 | 40 | 35 | 65 | 63 | 63 | 60 | 56 | 57 | 61 | 52 | 55 |
| OrdenezA | 53 | 56 | 54 | 77 | 68 | 72 | 75 | 71 | 72 | 73 | 69 | 71 |
| OrdenezC | 66 | 66 | 65 | 75 | 69 | 72 | 73 | 65 | 69 | 75 | 70 | 72 |
| Promedio | 49 | 55 | 51 | 71 | 65 | 67 | 68 | 62 | 64 | 69 | 64 | 66 |

También es reseñable el hecho de que hibridar un clasificador con un HMM introduce una mayor complejidad al sistema, por lo que debe tenerse en cuenta en entornos donde la eficiencia prime sobre la precisión. No obstante, dicha arquitectura híbrida ofrece una gran flexibilidad a la hora de emplear complejos y diversos espacios de características, a diferencia de los HMM, cuya necesidad de definir explícitamente las distribuciones que modelan las observaciones supone una limitación. Por ejemplo, nuestras propuestas híbridas podrían aceptar que se integrasen sensores no binarios, o de otra naturaleza, en la red de monitorización de una forma natural, sin requerir cambios en el modelo, al contrario que los HMM.

Además, la fase de entrenamiento de nuestro sistema no se hace en tiempo real, por lo que no consideramos el aumento de complejidad en el modelo híbrido como un problema serio. Sin embargo, este modelo sí plantea el problema de requerir conjuntos de datos totalmente etiquetados, a diferencia de los modelos generativos, que pueden llevar a cabo un aprendizaje semisupervisado si así se requiere.

Capítulo 5

Detección de Anomalías

En el presente capítulo detallamos nuestra propuesta para la detección automática de anomalías en el comportamiento de una persona, a través de una red de monitorización compuesta por sensores binarios. El objetivo de identificar patrones extraños en el comportamiento es poder proporcionar una ayuda a las personas con algún tipo de discapacidad o trastorno, que deseen vivir de forma independiente en la comodidad de su propio hogar. Presentamos las tres características que proponemos para modelar los patrones de comportamiento del usuario, y el método estadístico no supervisado encargado de la detección de valores atípicos.

La detección de anomalías es un problema relevante y una línea de investigación muy activa en diversas áreas y dominios de aplicación. Con el nombre de *detección de anomalías* se hace referencia a la problemática de encontrar patrones sobre una serie de datos, los cuales no se ajusten con un comportamiento esperado. A este tipo patrones se les denomina como anomalías, valores atípicos, observaciones discordantes, excepciones o aberraciones, dependiendo del tipo de dominio sobre el se trabaje. Aunque, a menudo, las denominaciones anomalías y valores atípicos son las más comunes, y las que vamos a emplear en esta tesis doctoral.

La detección de anomalías tiene un amplio abanico de aplicaciones, tales como la detección de fraudes, la seguridad informática o el cuidado de personas. Dependiendo del dominio, la información que ofrecen estos métodos puede llegar a ser muy importante, o incluso crítica. Por ejemplo, un tráfico de datos extraño dentro de una red privada de ordenadores puede significar que la seguridad de la instalación tiene algún fallo, o la aparición de valores atípicos en transacciones comerciales puede ser debida a un robo de identidad.

Una anomalía se define como una observación (o grupo de observaciones) que aparenta ser incoherente con el resto de observaciones pertenecientes al conjunto de datos y, por lo tanto, no se adhiere a la definición de valor normal. El enfoque más inmediato para desarrollar un sistema de detección de anomalías es mediante la

definición de una región en el espacio de características del problema que represente a los valores normales, y donde todo ejemplo que no pertenezca a esa región sea considerado como una anomalía. Pero este problema aparentemente sencillo plantea varios desafíos:

- El definir una región en el espacio que sea capaz de englobar a los valores no anómalos es una tarea complicada. Además, el límite que separa a una observación anómala de una normal es a menudo poco preciso. Especialmente complicado es definir la naturaleza de aquellas observaciones que se ubiquen cerca de dichos límites.
- En escenarios donde el comportamiento normal sea variable y se modifique a lo largo del tiempo, el sistema debe ser capaz de adaptarse a dichos cambios.
- El concepto de qué es una anomalía es diferente para cada dominio de aplicación. Por lo que aplicar una técnica útil para un dominio sobre otro dominio diferente no es algo sencillo.
- La disponibilidad de conjuntos de datos etiquetados influye de forma considerable en la usabilidad de las diferentes técnicas de detección de anomalías.
- Dentro de los conjuntos de datos puede existir ruido, el cual suele confundirse con datos atípicos.

Estos factores convierten a la detección de anomalías en una cuestión no trivial, y también representan la base que se emplea para clasificar los distintos métodos y técnicas.

Cuando se hace mención a que un conjunto de estas características está etiquetado, significa que las instancias de datos están previamente divididos en normales y anómalas. Pero debe tenerse en cuenta que obtener un conjunto de datos que contenga un etiquetado preciso de todos los posibles tipos de comportamientos en el dominio de aplicación es generalmente prohibitivamente costoso, ya que debe llevarse a cabo por un experto humano.

En función de la disponibilidad de datos etiquetados, las técnicas de detección de anomalías se dividen en supervisadas, semisupervisadas y no supervisadas. Las técnicas supervisadas trabajan con datos que están divididos en dos clases, normales y anómalos, dando lugar por tanto a un problema de clasificación binaria. Un problema típico en la detección supervisada de anomalías es el desbalanceo de los datos, puesto que el número de valores atípicos suele ser muy pequeño en comparación con el resto del conjunto [41].

Las técnicas de detección semisupervisadas asumen que en el conjunto de datos las instancias etiquetadas se corresponden con los datos no anómalos, no requiriendo una etiqueta para los valores atípicos. Por ejemplo, un caso de detección semisupervisada de anomalías es el propuesto por Fujimaki *et. al.*, donde emplearon técnicas de aprendizaje automático para modelar el comportamiento habitual de aeronaves, y de ese modo identificar escenarios donde podrían darse situaciones de accidente [81].

Las técnicas de detección no supervisada no requieren de la existencia de datos etiquetados en el conjunto, por tanto son las más comúnmente empleadas en la literatura, y el enfoque que tomamos para la propuesta que presentamos en este capítulo. Estas técnicas se basan en la premisa de que los valores no anómalos son aquellos más comunes dentro del conjunto de datos.

Un importante aspecto de estas técnicas es la manera en que informan de la detección de un valor atípico. La salida que ofrecen estos métodos puede ser normalmente de dos tipos, mediante etiquetas o mediante probabilidades. La asignación de una etiqueta simplemente consiste en indicar si cada instancia de prueba pertenece al conjunto de valores normales o al de anomalías. Para las técnicas que ofrecen como resultado una probabilidad, ésta representa el grado de confianza de que dicho ejemplo de prueba pertenezca al conjunto de los valores atípicos. Obtener una probabilidad como salida del método es la solución más versátil, puesto que al ordenar dichas probabilidades mediante un ranking, la diferenciación entre valores normales y atípicos se puede llevar a cabo estableciendo un límite, y no solo eso, sino que se pueden identificar los anomalías más relevantes. Es por ello que éste es el tipo de salida que ofrece nuestra propuesta.

En relación al mecanismo empleado para la identificación de anomalías existen diversas propuestas basadas en diferentes áreas de investigación. Un método muy presente en la literatura es el uso de clasificadores que lleven a cabo una generalización sobre un conjunto de entrenamiento para después realizar una clasificación o etiquetado sobre el conjunto de test. Un modelo capaz de distinguir entre valores normales y atípicos debe ser generado dado el espacio de características del problema. Algoritmos bien conocidos de este área como son las RNA [232], las redes bayesianas [255], las SVM [143] o los algoritmos basados en reglas [227] han sido aplicados con éxito a este problema, aunque un inconveniente de este tipo de enfoques es que no pueden indicar mediante ninguna métrica el grado de confianza de la clasificación.

Otros autores han presentado propuestas empleando el concepto de clasificación mediante el *vecino más cercano*. Estas técnicas se basan en la idea de que los valores no anómalos se encuentran en las zonas más densamente pobladas del espacio de datos, mientras que las anomalías, al ser valores extraños, se alejan de dichas áreas. El empleo de este concepto requiere definir algún tipo de formulación que sirva para determinar la distancia o similaridad existente entre dos

instancias cualesquiera. Es común el empleo de la distancia euclídea como medida de similitud para calcular si un valor es atípico, tomando como referencia los k valores más cercanos para llevar a cabo la clasificación [266, 4]. Este modo de detección tiene la ventaja de poder ser aplicado a problemas no supervisados, sin embargo, si la distribución espacial de los datos no es adecuada, y los valores atípicos se encuentran demasiado cerca de los normales, el método del *vecino más cercano* resulta muy ineficiente.

Otro método de detección presente en la literatura es el agrupamiento o *clustering*, una técnica principalmente no supervisada cuyo objetivo es agrupar ejemplos similares en conjuntos. Este método asume que los valores normales del conjunto de datos pertenecen a un mismo grupo, y cualquier ejemplo que encaje en la definición de ese grupo será considerado como un valor anómalo. Para esto es común el empleo de un centroide, esto es, un valor que funciona como centro de masas del conjunto y que sirve como referencia para estimar la distancia entre cada grupo y los nuevos valores. Diferentes técnicas de detección de anomalías han sido propuestas tomando como base este sistema [244, 33]. De entre las diferentes técnicas para la detección de anomalías, posiblemente la más empleada, y la que se ha elegido para esta propuesta, es el enfoque estadístico. Esta técnica se basa en el principio de que una anomalía es toda aquella observación que se considere parcial o totalmente irrelevante por un modelo estocástico previamente definido [5]. Por tanto, los valores normales del conjunto son aquellos que aparecen en las regiones de alta probabilidad del modelo estocástico. Mediante este enfoque se ajusta un modelo estadístico (por lo general definitorio del comportamiento normal) a los datos, para después aplicar inferencia estadística sobre las nuevas instancias y así determinar si pertenecen al modelo (y por lo tanto no son atípicos) o no. Estas técnicas suelen ser de tipo paramétrico, y requieren definir las distribuciones que modelarán el espacio de características. En la literatura existen ejemplos de detección estadística de anomalías basados en distribuciones gaussianas [12], regresión lineal [43] y mixturas de distribuciones [2]. Este modo de detección tiene las ventajas de ofrecer como salida una probabilidad (que es interpretable como un nivel de confianza) y puede ser aplicado a problemas no supervisados. Aunque su principal inconveniente es precisamente que estas técnicas se basan en la suposición de que los datos se generan a partir de una distribución en particular.

El enfoque de este trabajo está orientado a desarrollar un modelo que sea capaz de detectar patrones anómalos en el comportamiento de una persona, para eventualmente poder identificar y llegar a predecir situaciones de riesgo para la salud, con el objetivo de promover la independencia de personas con trastornos físicos o cognitivos. Poder llegar a comprender el comportamiento humano tradicionalmente ha despertado mucho interés en la comunidad investigadora, y se ha dedicado mucho trabajo a proponer soluciones para el modelado del comportamiento, con el objetivo de desarrollar sistemas que puedan considerar en algún

grado el comportamiento de los usuarios. Mozer *et. al.* [170] y Chang *et. al.* [38] presentaron algunas de las primeras aplicaciones donde los entornos monitorizados tenían en cuenta los patrones de comportamiento de los usuarios. Mediante la observación de los movimientos de las personas, estos modelos realizaban predicciones acerca de la posible ocupación de las diferentes estancias de una casa.

Si una vivienda puede ser monitorizada mediante uno de estos sistemas, los habitantes tendrán mejores opciones para poder vivir de forma más segura e independiente [184], especialmente si estas personas sufren de alguna afección crónica o discapacitante, como pueden ser la enfermedad de Alzheimer o Parkinson. Diversos estudios se han enfocado especialmente a integrar estos modelos de detección con sistemas de alarmas, con el objetivo de poder alertar a familiares o cuidadores en el caso de que se detecten comportamientos extraños en los usuarios [251].

En la literatura existen diversos trabajos orientados a detectar desviaciones o acciones extrañas en el comportamiento de los usuarios, mediante diferentes tecnologías de monitorización. Recientemente, Aztiria *et. al.* desarrollaron un algoritmo para comparar el comportamiento del usuario con un conjunto de patrones frecuentes, previamente aprendidos, con el fin de identificar desviaciones [9]. Mediante la definición de un conjunto de posibles acciones atómicas, estos autores calculan una probabilidad para el comportamiento del usuario en un determinado momento, y así poder detectar si es un comportamiento habitual. El número de modificaciones requeridas para transformar el comportamiento actual en uno de los que son considerados como habituales es empleado como métrica para determinar si una secuencia de acciones concreta debe ser considerada como una anomalía.

Shin *et. al.* presentaron un sistema de detección de anomalías que capturaba información sobre el nivel de actividad y movilidad de los usuarios para identificar patrones de comportamiento atípicos [216]. Estos autores propusieron un sistema diseñado para monitorizar la vida diaria de una persona mayor en su propio domicilio mediante el uso de varios sensores infrarrojos de movimiento. Para poder representar el comportamiento de los usuarios del sistema, propusieron el empleo de tres características, a saber: nivel de actividad, nivel de movilidad e intervalo sin respuesta. Un método de clasificación binaria basado en un algoritmo de vectores de soporte se empleó para combinar la información obtenida a través de dichas características y para identificar los patrones atípicos. Los autores emplearon para la validación de su propuesta datos obtenidos de entornos reales y de entornos simulados. Este método obtuvo una sensibilidad del 74.2 % y una especificidad del 85.8 % para los conjuntos de datos reales, e incluso un rendimiento mayor para los valores simulados.

Yin *et. al.* también emplearon una SVM entrenada para clasificación binaria con el objetivo de detectar acciones atípicas mediante el uso de sensores portá-

tiles [263]. Su enfoque empleaba una cadena de procesos conectados en tubería que combinaba una SVM para filtrar los eventos normales con una colección de clasificadores secundarios para detectar los datos anómalos. Compararon tres versiones diferentes de su propuesta en términos de área bajo la curva ROC (AUC), concluyendo que su propuesta obtenía un mejor rendimiento en términos generales.

Otro estudio basado en SVMs fue presentado por Palaniappan *et. al.* [181]. En este caso, los autores emplearon una SVM multiclase para reconocer los acciones normales, y los patrones atípicos fueron detectados descartando todas las posibles acciones que podrían llevarse a cabo a continuación. El diseño de este método estaba enfocado a detectar actividades anómalas con el menor coste computacional posible, para poder ser explotado en tiempo real. De hecho, las SVMs han sido empleadas para detectar comportamientos anómalos no solamente con datos obtenidos de WSNs, sino también se han integrado con videocámaras [257]. Wu *et. al.* definieron un método que usaba una transformada discreta de Fourier para obtener diversas características de una secuencia de datos de vídeo, y mediante una SVM, clasificar los patrones en normales o atípicos.

Otras propuestas para modelar los patrones de comportamiento de los usuarios han tomado como base los ritmos circadianos (*Circadian Activity Rhythms* o CAR en inglés). Virone *et. al.* desarrollaron un sistema de monitorización con una clara orientación a fomentar la independencia de las personas mayores que viven solas. El objetivo de este sistema era capturar patrones de comportamiento mediante el uso de algoritmos estadísticos que pudieran modelar los ritmos circadianos y sus posibles desviaciones [247]. Los autores presentaron un sistema compuesto por diferentes tipos de sensores inalámbricos para monitorizar el estado de salud y el comportamiento de una persona dentro de su propio hogar [246]. Dicho sistema analizaba diferentes métricas estadísticas, mediante el nivel de actividad y presencia del usuario, para detectar desviaciones en los patrones del comportamiento. Para poder diferenciar los comportamientos no habituales definieron un intervalo de credibilidad.

El comportamiento humano es por naturaleza no determinista, por lo que definir un método que pueda modelarlo de forma precisa no es una cuestión trivial. No obstante, existen propuestas, tales como los métodos bayesianos, que ofrecen una buena base para el desarrollo de técnicas de modelado que específicamente tengan en cuenta la incertidumbre de los datos. De hecho, la estadística bayesiana ha demostrado ser particularmente apropiada para el modelado del comportamiento humano [74] y la detección de anomalías [107].

En este capítulo presentamos un modelo estadístico no supervisado para llevar cabo la detección automática de patrones anómalos. En esta tesis doctoral postulamos que es posible obtener una estimación precisa del comportamiento de una persona estadísticamente mediante modelos bayesianos. Para ello propo-

nemos tres valores probabilísticos para modelar el comportamiento del usuario, que son estimados a través de los datos de los sensores binarios, y hasta donde nosotros sabemos, no han sido utilizados en la literatura.

En el siguiente apartado se explica la metodología de nuestra propuesta para la detección de valores atípicos, detallando los valores probabilísticos presentados. A continuación, describimos la configuración experimental y los resultados del modelo. Por último, finalizamos el capítulo con las conclusiones.

5.1. Metodología

En este apartado definimos la base teórica de nuestro método para identificar anomalías en el comportamiento de una persona, en base a tres métricas estadísticas que obtenemos de los datos capturados a través de nuestra red de sensores inalámbricos.

Postulamos que es posible modelar los patrones de comportamiento del usuario empleando probabilidades definitorias de: (1) cuándo se activan los sensores, (2) en qué orden y (3) durante cuánto tiempo.

Estas tres probabilidades ofrecen información relevante con la que poder valorar la anormalidad de las acciones y los patrones de una persona. Dichas probabilidades las hemos denominado como:

1. Probabilidad de activación de los sensores (PAS) - Esta probabilidad es representativa de los ritmos circadianos del usuario, y por tanto, tiene un importante valor potencial para la monitorización de una persona a largo plazo.
2. Probabilidad de secuencia de los sensores (PSS) - Esta probabilidad modela el orden de activación de los sensores, y puede resultar útil en la identificación de situaciones de confusión o delirio.
3. Probabilidad de duración de los sensores (PDS). Esta probabilidad ofrece información acerca de cuanto tiempo se mantiene activo un sensor. Es representativa de la condición física del usuario, resultando útil en la identificación de situaciones de caídas o desmayos.

Además, todas ellas pueden resultar útiles a la hora de determinar si el usuario padece algún trastorno o desorden cognitivo. Cualquier de estas métricas por sí sola ofrece información útil para evaluar el comportamiento, sin embargo nosotros también consideramos combinar dichas informaciones, debido a que es muy probable que la precisión en la detección de anomalías mejore.

Dichas probabilidades son estimadas mediante estadística bayesiana, puesto que este enfoque se adapta mejor a nuestro planteamiento del problema, en el

| Métricas | Distribuciones | | Distribuciones a priori | |
|----------------------------|----------------|----------------|-------------------------|----------------------|
| | Nombre | Parámetros | Nombre | Hiperparámetros |
| Activación de los sensores | Bernoulli | ρ | Beta | α, β |
| Secuencia de los sensores | Multinomial | θ | Dirichlet | η |
| Duración de los sensores | Normal | μ, λ | Normal-Gamma | ϕ, κ, a, b |

Tabla 5.1: Resumen de las distribuciones empleadas para modelar las características, y sus correspondientes distribuciones a priori.

sentido que en el análisis bayesiano la probabilidad es interpretada como la representación del grado de creencia en una hipótesis [22].

Abordar nuestro problema a través de la estadística frecuentista, interpretando la probabilidad mediante la frecuencia de los eventos, no es una solución adecuada a nuestro planteamiento, puesto que la distribución exacta de las probabilidades que proponemos será siempre desconocida (ya que nunca dispondremos de infinitos datos para aprender las distribuciones). En cambio, el análisis bayesiano nos permite estimar la esperanza de nuestras distribuciones, mediante la marginalizando de los parámetros. Por ejemplo, estimar la probabilidad de que el usuario active el sensor de la ducha a las ocho de la mañana, mediante el método de simplemente contar cuantas veces ha ocurrido así en los últimos días, nos proporciona una información muy simple y limitada de la probabilidad de activación de ese sensor, sin embargo, la estadística bayesiana nos ofrece además una estimación de ese suceso a priori, tal y como se ha mostrado anteriormente en esta tesis mediante el ejemplo de la moneda, en el Apartado 4.4.

Por esto, modelamos las métricas propuestas mediante distribuciones a priori, las cuales representan el grado de conocimiento inicial que tenemos de los valores de los parámetros (y por tanto del comportamiento de la persona). Al igual que en la solución propuesta en el Apartado 4.4.1 para el problema de reconocimiento de actividades, empleamos las distribuciones conjugadas como distribuciones a priori, puesto que son particularmente adecuadas. Como recordatorio, se entiende por familia conjugada de distribuciones a priori, a aquellas distribuciones que cuando son combinadas mediante el teorema de Bayes, resultan en una distribución a posteriori que es de la misma familia paramétrica de la distribución a priori. Si disponemos de una adecuada información a priori, el análisis bayesiano ofrece resultados más precisos y menos sesgados que la estimación por máxima verosimilitud.

Tal y como se ha explicado en el anterior capítulo de esta tesis doctoral, los parámetros que gobiernan las distribuciones a priori son denominados hiperparámetros. Un resumen de las diferentes distribuciones de las métricas y sus correspondientes parámetros e hiperparámetros puede verse en la Tabla 5.1.

5.1.1. Probabilidad de Activación de los Sensores : PAS

Con este valor, se representa la probabilidad de que un sensor en particular se active durante un intervalo de tiempo concreto. Para ello, se requiere que el flujo de datos sea segmentado en intervalos regulares, tal y como se propuso en el Apartado 3.3 de esta tesis doctoral. Empleando ese mismo preprocesado de datos se obtiene un vector de valores binarios, que representa las activaciones de los sensores, para cada intervalo de tiempo. Esto es, los eventos de sensores se expresan como x_t^i , indicando si el sensor i se ha activado al menos una vez durante el intervalo definido entre el instante t y el instante $t + \Delta t$, donde $x_t^i \in \{0, 1\}$ y la longitud de los intervalos está definida como Δt .

La activación de cada sensor en un determinado intervalo está modelada mediante una distribución de Bernoulli, donde ρ_t^i es el parámetro del sensor i para el intervalo t . Por tanto, la probabilidad de que el usuario active el sensor i durante el intervalo t se define mediante la ecuación:

$$p(x_t^i | \rho_t^i) = (\rho_t^i)^{x_t^i} (1 - \rho_t^i)^{1-x_t^i} \quad (5.1)$$

En nuestra propuesta, hemos adoptado una solución para estimar el parámetro ρ_t^i que toma en cuenta no sólo el intervalo actual t , sino también los valores del sensor en un determinado número de intervalos alrededor de t . No tiene demasiado sentido intentar estimar el comportamiento de un sensor en un intervalo específico, cuando la actividad de ese mismo sensor un instante antes o un instante después está siendo ignorada. Es por ello que definimos un parámetro W que proporciona un mecanismo simplificado de ventana deslizante para la estimación del parámetro ρ_t^i . El parámetro W define la dimensión de la ventana $\langle t - W/2, \dots, t, \dots, t + W/2 \rangle$ que se empleará en el cálculo de ρ_t^i .

Debido a que para estimar esta probabilidad empleamos una distribución de Bernoulli, su distribución conjugada natural tiene la forma de una distribución beta, definida en la ecuación 5.2.

$$p(\rho_t^i | \alpha_t^i, \beta_t^i) = \text{Beta}(\rho_t^i | \alpha_t^i, \beta_t^i) = \frac{1}{B(\alpha_t^i, \beta_t^i)} \rho_t^{i\alpha_t^i-1} (1 - \rho_t^i)^{\beta_t^i-1} \quad (5.2)$$

Donde $B(\alpha_t^i, \beta_t^i)$ es la función beta. Por lo tanto, la probabilidad a posteriori para el parámetro ρ_t^i se define en la ecuación 5.3.

$$p(\rho_t^i | x_{\{t-W/2\}}^i \dots x_{\{t+W/2\}}^i) \propto p(x_{\{t-W/2\}}^i \dots x_{\{t+W/2\}}^i | \rho_t^i) \cdot (\rho_t^i) \propto \text{Beta}(\rho_t^i | \alpha_t^i + R_t^i, \beta_t^i + S_t^i) \quad (5.3)$$

Donde las ecuaciones 5.4 y 5.5 representan, respectivamente, el número de observaciones de unos y ceros para el sensor i dentro de la ventana W centrada en t .

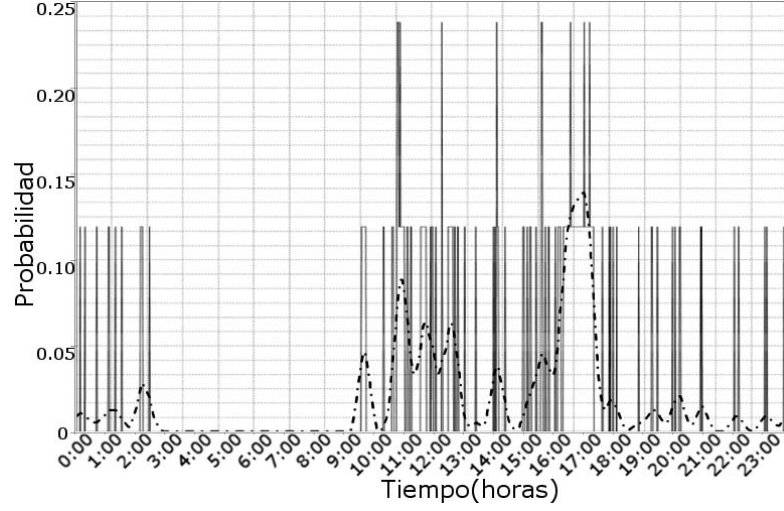


Figura 5.1: Probabilidad de activación de un sensor instalado en el lavabo empleando estadística frecuentista (línea continua) y nuestra probabilidad (PAS) (línea punteada).

$$R_t^i = \sum_{z=t-W/2}^{t+W/2} r_z^i \xi \quad (5.4)$$

$$S_t^i = \sum_{z=t-W/2}^{t+W/2} s_z^i \xi \quad (5.5)$$

Siendo ξ un factor de ponderación para ajustar la relevancia relativa de la ventana sobre el valor final. Las ecuaciones 5.6 y 5.7 representan, respectivamente, el número de observaciones de unos y ceros para el sensor i en el intervalo t durante el número D de días de los que se compone el conjunto de datos; $\mathbb{I}_{\{\cdot\}}$ denota a la función indicatriz.

$$r_t^i = \frac{1}{D} \sum_{d=1}^D \mathbb{I}_{\{x_t^i=1\}} \quad (5.6)$$

$$s_t^i = \frac{1}{D} \sum_{d=1}^D \mathbb{I}_{\{x_t^i=0\}} \quad (5.7)$$

Al comenzar la estimación de la métrica se deben seleccionar unas distribuciones a priori no informativas que no puedan influir sobre el cálculo de la probabilidad. Como valores iniciales para los parámetros de la distribución a priori de la PAS empleamos $\alpha_0^i = \beta_0^i = 1$, dando lugar a unos valores a priori suficientemente difusos. Como paso final, mediante la integración de los parámetros del modelo,

estimamos la probabilidad de activación del sensor i en el intervalo t según la ecuación 5.8.

$$\begin{aligned}
 \text{PAS}_t(i) &= p(x_t^i = 1 | x_{\{t-W/2\}}^i \dots x_{\{t+W/2\}}^i) = \\
 &\int_0^1 p(x_t^i = 1 | \rho_t^i) p(\rho_t^i | x_{\{t-W/2\}}^i \dots x_{\{t+W/2\}}^i) d\rho_t^i = \\
 &\int_0^1 \rho_t^i \text{Beta}(\rho_t^i | \alpha_t^i + R_t^i, \beta_t^i + S_t^i) d\rho_t^i = \\
 E[\rho_t^i] &= E[\rho_t^i] = \frac{(\alpha_t^i + R_t^i)}{(\alpha_t^i + R_t^i) + (\beta_t^i + S_t^i)}
 \end{aligned} \tag{5.8}$$

En la Figura 5.1 se muestra una representación de esta probabilidad a lo largo de un día, donde $\Delta t = 60$ segundos y $W = 50$ (ventana de 50 minutos).

5.1.2. Probabilidad de Secuencia de los Sensores : PSS

El objetivo de estimar la probabilidad de secuencia es poder modelar el orden de activación de los sensores, con la finalidad de inferir la estructura del plan de acción más plausible para el usuario.

Para estimar esta métrica se toma como entrada el orden de activación de los sensores. El primer paso es interpretar la activación de los sensores como una secuencia de eventos, representada por S , a partir del conjunto de datos. La probabilidad de transición $p(s_j | s_i)$ representa la probabilidad de que el sensor s_j se active inmediatamente después del sensor s_i . Estas probabilidades se definen mediante N distribuciones multinomiales, una para cada sensor, donde las probabilidades individuales de transición se representan como $\theta_{ij} \equiv p(s_t = j | s_{t-1} = i)$.

Para este caso, al emplear una distribución multinomial, su distribución conjugada natural tiene la forma de una distribución de Dirichlet, definida en la ecuación 5.9.

$$p(\vec{\theta}_i | \vec{\eta}_i) = \text{Dir}(\vec{\theta}_i | \vec{\eta}_i) = \frac{1}{Z(\vec{\eta}_i)} \cdot \theta_{i1}^{\eta_{i1}-1} \dots \theta_{iN}^{\eta_{iN}-1} \tag{5.9}$$

donde

$$Z_{\text{Dir}}(\vec{\eta}_i) = \frac{\prod_{j=1}^N \Gamma(\eta_{ij})}{\Gamma(\sum_{j=1}^N \eta_{ij})} \tag{5.10}$$

y la suma de las probabilidades individuales es igual a uno ($\sum_k \theta_{ik} = 1$). Para estimar esta métrica, elegimos como valores iniciales, no informativos, de los parámetros de la distribución a priori $\forall \eta_0 = 1$. Por tanto, la probabilidad a posteriori

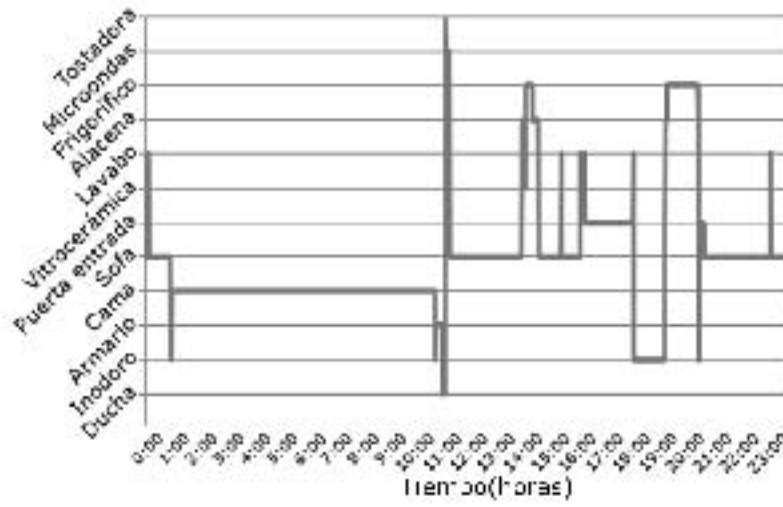


Figura 5.2: Secuencia S de activaciones de los sensores durante un día de datos para el conjunto ‘OrdenezA’.

para el parámetro $\vec{\theta}_i$, que representa la distribución multinomial del sensor i , se define en la ecuación 5.11.

$$p(\vec{\theta}_i | S, \vec{\eta}_i) = \text{Dir}(\eta_{i1} + U_{i1}, \dots, \eta_{iN} + U_{iN}) \quad (5.11)$$

donde

$$U_{ij} = \sum_{s=2}^S \mathbb{I}_{\{s_{t-1}=i, s_t=j\}} \quad (5.12)$$

es el número de transiciones desde el sensor i al sensor j en la secuencia de activaciones. Por último, calculamos la probabilidad de que el usuario active los sensores en un orden concreto, estimando la probabilidad de transitar desde el sensor s_i al sensor s_j como se define en la ecuación 5.13.

$$\text{PSS}_i(j) = p(\theta_{ij} | S) = \frac{\eta_{ij} + U_{ij}}{U_i + \sum_N \eta_{iN}} \quad (5.13)$$

En la Figura 5.2 se muestra un ejemplo de secuencia de activación de los sensores a lo largo de un día, con el tratamiento de datos propuesto.

5.1.3. Probabilidad de Duración de los Sensores : PDS

Esta es una métrica que refleja la duración de las actividades y movimientos del usuario, una información considerada relevante para la detección de anomalías

en el comportamiento. El primer paso en la obtención de la probabilidad es procesar la información de los sensores para obtener la duración de las activaciones de cada sensor, la cual viene definida en función de un número de intervalos. Al igual que en el cálculo de la probabilidad de activación de los sensores, el flujo de datos está discretizado en un conjunto de intervalos, cuya longitud está definida como Δt .

Para cada sensor i existe un conjunto $L_i := \{l_{i1} \dots l_{iN}\}$ compuesto por N valores, que representan las duraciones (en número de intervalos) de activación para ese sensor, siendo \bar{l} el promedio del conjunto. La duración de cada sensor se modela mediante una distribución normal, donde μ_i y λ_i son la media y la precisión¹, respectivamente, para la duración del sensor i . Emplear este tipo de distribución para representar la duración de las acciones humanas es una solución muy común en la literatura [249]. Dicha distribución se define como indica la ecuación 5.14.

$$p(L_i|\mu_i, \lambda_i) = \frac{1}{(2\pi)^{N/2}} \lambda_i^{N/2} \exp\left(-\frac{\lambda_i}{2} \sum_{n=1}^N (l_{in} - \mu_i)^2\right) \quad (5.14)$$

En este caso, empleamos una distribución normal-gamma como distribución a priori para los parámetros μ_i y λ_i , por lo que la distribución a priori de la probabilidad de duración de los sensores se define tal y como se indica en la ecuación 5.15.

$$NG(\mu_i, \lambda_i|\phi_i, \kappa_i, a_i, b_i) = \mathcal{N}(\mu_i|\phi_i, (\kappa_i \lambda_i)^{-1}) Ga(\lambda_i|a_i, b_i) = \frac{1}{Z_{NG}} \lambda_i^{a_i - \frac{1}{2}} \exp\left(-\frac{\lambda_i}{2} [\kappa_i (\mu_i - \phi_i)^2 + 2b_i]\right) \quad (5.15)$$

donde

$$Z_{NG}(\phi_i, \kappa_i, a_i, b_i) = \frac{\Gamma(a_i)}{b_i^{a_i}} \left(\frac{2\pi}{\kappa_i}\right)^{\frac{1}{2}} \quad (5.16)$$

Empleamos los valores $\kappa_i = a_i = b_i > 0,1$ y $\phi_i = 0$, dando como resultado una distribución a priori no informativa. Al igual que otros autores [124], no hemos encontrado diferencia en los resultados cuando comparamos nuestra distribución a priori normal-gamma con otras propuestas conjugadas no informativas. Por tanto, la probabilidad a posteriori para los parámetros μ_i y λ_i , que modelan la distribución normal de probabilidad para la duración del sensor i se representa mediante la ecuación 5.17.

$$p(\mu_i, \lambda_i|L_i) = NG(\mu_i, \lambda_i|\phi'_i, \kappa'_i, a'_i, b'_i) \quad (5.17)$$

¹la precisión en una distribución normal se define como 1/varianza.

donde

$$\phi'_i = \frac{\kappa_i \mu_i + N \bar{l}}{\kappa_i + N} \quad (5.18)$$

$$\kappa'_i = \kappa_i + N \quad (5.19)$$

$$a'_i = a_i + N/2 \quad (5.20)$$

$$b'_i = b_i + \frac{1}{2} \sum_{n=1}^N (l_n - \bar{l})^2 + \frac{\kappa_i N (\bar{l} - \mu_i)^2}{2(\kappa_i + N)} \quad (5.21)$$

Por último, podemos estimar la probabilidad de que el sensor i esté activo durante t intervalos, mediante una distribución t de Student, tal y como se indica en la ecuación 5.22.

$$\text{PDS}_i(1) = p(l|L_i) = t_{2a'_i}(l|\phi'_i, \frac{b'_i(\kappa'_i + 1)}{a'_i \kappa'_i}) \quad (5.22)$$

5.1.4. Clasificación de las Anomalías

Para estimar si la salida de alguna de las métricas propuestas debe ser considerada atípica o no, hacemos uso de un intervalo de credibilidad bayesiano (BCI por sus siglas en inglés). En el análisis bayesiano, los intervalos de credibilidad son la contraparte del concepto de los intervalos de confianza en el análisis estadístico clásico. Un intervalo de credibilidad de $100(1 - \alpha)\%$ para $\theta \in \Theta$ es un subconjunto $C \subset \Theta$ tal que $p(\theta \in C|x) \geq 1 - \alpha$. O lo que es lo mismo, un intervalo de credibilidad del $100(1 - \alpha)\%$ para una proporción nos indica que hay una probabilidad igual al $100(1 - \alpha)\%$ de que la proporción de la población se encuentre en el intervalo.

Este mecanismo para estimar los valores atípicos ha sido empleado con éxito en otras propuestas de detección de anomalías presentes en la literatura [107]. Construyendo un BCI para cada métrica, las anomalías pueden ser detectadas secuencialmente a medida de que se dispone de nuevos valores de salida. Un BCI del $p\%$ indica que la probabilidad a posteriori de qué valor actual de la métrica caiga dentro del intervalo es p ; por ello, el BCI define el rango de valores admisibles para cada una de las métricas propuestas. Por lo tanto, cualquier valor que no pertenezca al $p\%$ es considerado una anomalía. Un BCI del $100(1 - \alpha)\%$ para un nuevo valor puede ser calculado mediante la ecuación 5.23.

$$\bar{x} \pm z_{\alpha/2} \sqrt{\sigma^2} \quad (5.23)$$

Donde $z_{\alpha/2}$ es el $(100(1 - \alpha/2))$ -ésimo percentil de una distribución normal, y σ^2 es la varianza de la distribución posterior marginal de la predicción de los valores. Debe tenerse en cuenta que para algunos casos (por ejemplo en el caso de la probabilidad de secuencia de los sensores) los valores considerados como anomalías son únicamente aquellos por debajo del rango de valores admisibles (la transición entre dos sensores es atípica sólo cuando es muy poco probable, no al contrario).

5.2. Evaluación

En este apartado presentamos los experimentos que hemos llevado a cabo para evaluar nuestro método de detección de anomalías. Describimos los conjuntos de datos, la configuración de los experimentos y los resultados de las pruebas. El primer experimento está enfocado a evaluar de forma individual el rendimiento de cada una de las métricas propuestas, con el objetivo de determinar cuál es el BCI más adecuado para cada una. En el segundo experimento se analiza el rendimiento global del sistema cuando se combinan las salidas de todas las métricas, y además se incluye una comparativa con otra propuesta del área, que ha ofrecido buenos resultados.

5.2.1. Conjuntos de Datos

Para llevar a cabo la validación de la propuesta, se han empleado conjuntos de datos con dos tipos de anomalías, reales y sintéticas. Para que el sistema presentado pueda ser considerado fiable, la metodología propuesta debe ser validada con datos reales. Sin embargo, debido a lo complejo que es detectar correctamente todos los eventos anómalos que puedan existir en dichos datos reales, también se incluyen errores sintéticos en los conjuntos de datos empleados en la evaluación, tal y como ha sido propuesto por diversos autores [216, 107], para aumentar la validación del sistema. Por lo tanto, ambos tipos de datos, tanto reales como sintéticos, han sido empleados para evaluar el rendimiento del método en la detección de situaciones anómalas.

Los datos reales son aquellos que fueron directamente capturados por nuestras redes de sensores. Estos datos necesitaron ser analizados de forma manual y los eventos atípicos identificados por los investigadores. Los conjuntos de datos sintéticos han sido generados tomando como base los conjuntos de datos reales, para que puedan ser incluidos en una comparativa y sirvan de complemento en la evaluación. Para cada conjunto de datos real, los errores sintéticos se han generado de acuerdo a la ecuación 5.24.

$$D^* = D \pm \Delta \quad (5.24)$$

Donde D^* es el valor sintético, D es el valor original, y Δ es un factor de compensación. Usando este mecanismo, las anomalías sintéticas fueron introducidas en cada conjunto de datos, en nuestro caso, con una frecuencia del 5 %.

Por lo tanto, para cada uno de los conjuntos de datos descritos en la Tabla 3.1 se ha generado un conjunto de datos equivalente, pero con el añadido de los errores sintéticos, que han sido generados de forma artificial. Este mecanismo además es útil para verificar si el rendimiento de nuestra propuesta se incrementa asintóticamente a medida que se introducen más valores anómalos.

Debe tenerse en cuenta que los conjuntos de datos sintéticos no han sido creados a partir de cero, sino que son versiones modificadas de los conjuntos de datos que han sido generados empleando nuestros entornos de monitorización. Entonces, para cada uno de los conjuntos de datos reales, existe una versión sintética de ese conjunto, donde el número de anomalías es notablemente superior (contiene las anomalías reales más las anomalías sintéticas).

5.2.2. Configuración de la Experimentación

Para llevar a cabo la evaluación se han empleado los tres conjuntos de datos obtenidos a través de las tres redes de sensores desplegadas. Dichos conjuntos de datos son los descritos en la Tabla 3.1. Los datos capturados por los sensores cuando más de una persona estaba presente en la vivienda han sido descartados, debido a que tal y como está diseñado el sistema, no se pueden gestionar esas situaciones. Los perfiles de los diferentes usuarios son los detallados en el Apartado 3.1.3 de esta tesis doctoral, donde se muestra que la principal diferencia entre ellos es su nivel de autonomía.

Para dividir los datos en entrenamiento y test, en este caso también se empleó una validación cruzada del tipo ‘dejando uno fuera’ (*Leave-one-out cross-validation*). Para cada conjunto de datos, la validación cruzada se compone de tantas iteraciones como días tenga el conjunto, y en cada iteración, un día sirve como conjunto de test y el resto como conjunto de entrenamiento. Este proceso se repite para todas las iteraciones, empleando en cada iteración una submuestra diferente y obteniendo al final un resultado promedio.

Para poder estimar la PAS y la PDS el flujo de datos necesita estar discretizado en un conjunto de intervalos. Para el cálculo de estas probabilidades los datos se han segmentado empleando una granularidad $\Delta t = 60$ segundos.

| | | Inferido | |
|------|---------|-------------------------|-------------------------|
| | | Normal | Anómalo |
| Real | Normal | Verdadero Positivo (VP) | Falso Negativo (FN) |
| | Anómalo | Falso Positivo (FP) | Verdadero Negativo (VN) |

Tabla 5.2: Matriz de confusión

5.2.3. Optimización de los Parámetros

En el primer experimento se lleva a cabo un análisis individual de cada una de las métricas presentadas en nuestra propuesta. Para evaluar su rendimiento se hace uso tanto de los conjuntos de datos originales, como de los generados por nosotros mediante la inclusión de anomalías sintéticas. Las tres métricas son evaluadas individualmente.

Los conjuntos de datos empleados para la detección de anomalías típicamente sufren un fuerte desbalanceo debido a su naturaleza, siendo el número de elementos anómalos muy inferior al número de elementos normales. Por ello, para evaluar el rendimiento de las métricas se emplea la medida F o *F-Measure*, una medida ponderada que considera todas las clases igualmente importantes al combinar el Valor Predictivo Positivo (PPV) con la Razón de Verdaderos Positivos (VPR).

La detección de anomalías es un problema binario, por lo que su matriz de confusión se representa tal y como se indica en la Tabla 5.2. Las filas representan los valores reales de los conjuntos, mientras que las columnas indican los valores inferidos por nuestro sistema. Por lo tanto, los verdaderos positivos representan el número de eventos anómalos correctamente identificados; los falsos negativos el número de eventos anómalos sobre los que se ha cometido un error y no se han identificado; los falsos positivos el número de eventos normales que el sistema ha identificado erróneamente como atípicos y los verdaderos negativos el número de eventos normales que se han identificado correctamente como tales.

El *F-Measure* ya lo hemos definido anteriormente en el Apartado 4.3.2 de esta tesis doctoral, sin embargo, al cambiar ligeramente la matriz de confusión (este problema equivaldría a una clasificación binaria), ahora el Valor Predictivo Positivo (PPV) y la Razón de Verdaderos Positivos (VPR) se redefinen cómo se detalla en las ecuaciones 5.25 y 5.26.

$$\text{PPV} = \frac{VP}{VP + FP} \quad (5.25)$$

$$\text{VPR} = \frac{VP}{VP + FN} \quad (5.26)$$

En nuestra propuesta el BCI determina el equilibrio entre la proporción de eventos anómalos que son correctamente identificados y la proporción de eventos normales que son correctamente identificados como tales. El PPV representa la proporción de anomalías inferidas que realmente se corresponden con un comportamiento anormal, mientras que la VPR mide la proporción de anomalías reales que son correctamente identificadas.

En las Tablas 5.3, 5.4 y 5.5 se muestra el rendimiento de cada métrica propuesta para cada uno de los conjuntos de datos ('OrdonezA', 'OrdonezB' y 'OrdonezC', respectivamente) cuando son evaluadas usando BCIs del 80 %, 85 %, 90 %, 95 % y 98 %. Se puede apreciar cómo el PPV aumenta a medida que el BCI decrece, tal y como se podría esperar. En términos generales, la precisión del modelo mejora con el conjunto de datos sintético, debido a la incorporación de activaciones de los sensores inherentemente anómalas.

Se observa cómo el rendimiento general del modelo está influido muy significativamente por el valor del BCI. La métrica PAS obtiene de forma consistente los mejores resultados cuando se usa un BCI del 95 %, tanto para los datos reales como para los sintéticos. La PSS no obtiene de forma consistente los mejores resultados con un BCI concreto, aunque la precisión aumenta con un BCI alto y al añadir datos sintéticos, más concretamente, es capaz de reducir los casos de falsos negativos.

En el caso de la PDS, ésta obtiene los mejores valores con los conjuntos de datos reales cuando el BCI es alto, y al contrario que las otras métricas, su rendimiento empeora cuando se introducen errores sintéticos. Los mejores resultados se consiguen cuando se usa un BCI del 98 % para los datos reales y un BCI del 80 % para los datos sintéticos. Esta métrica se comporta de modo algo diferente al resto debido a que el número de anomalías de este tipo aparece en menor proporción en los conjuntos, esto es, para esta métrica existe un valor en el conjunto por cada activación de un sensor, mientras que para las otras métricas existe un valor por cada intervalo de tiempo.

5.2.4. Resultados

La principal contribución de este capítulo de la tesis es presentar tres métricas estadísticas para modelar los patrones diarios de una persona que sirvan para detectar anomalías en el comportamiento mediante un análisis bayesiano. El siguiente experimento está enfocado a evaluar el rendimiento global del sistema, cuando las estimaciones de cada métrica son combinadas como una única salida del modelo. En este experimento, para obtener una mejor validación y con fines comparativos, se emplean cuatro medidas diferentes para evaluar la precisión del modelo.

Como se ha indicado previamente, en este dominio los conjuntos de datos no

| | | Conjunto de datos ‘OrdonezA’ | | | | | | | | |
|-----------|-----|------------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | | PAS | | | PSS | | | PDS | | |
| | BCI | PPV | VPR | FMeasure | PPV | VPR | FMeasure | PPV | VPR | FMeasure |
| Real | 98 | 0,92 | 0,59 | 0,72 | 0,34 | 0,62 | 0,44 | 0,86 | 0,99 | 0,93 |
| | 95 | 0,72 | 0,99 | 0,83 | 0,29 | 0,62 | 0,40 | 0,71 | 0,99 | 0,83 |
| | 90 | 0,37 | 0,99 | 0,54 | 0,29 | 0,68 | 0,41 | 0,60 | 0,99 | 0,75 |
| | 85 | 0,25 | 0,99 | 0,40 | 0,28 | 0,81 | 0,42 | 0,43 | 0,99 | 0,60 |
| | 80 | 0,19 | 0,99 | 0,32 | 0,26 | 0,87 | 0,40 | 0,39 | 0,99 | 0,56 |
| Sintético | 98 | 0,96 | 0,56 | 0,71 | 0,41 | 0,71 | 0,52 | 0,92 | 0,16 | 0,27 |
| | 95 | 0,89 | 0,89 | 0,89 | 0,32 | 0,24 | 0,43 | 0,92 | 0,25 | 0,40 |
| | 90 | 0,77 | 0,90 | 0,83 | 0,33 | 0,78 | 0,46 | 0,85 | 0,32 | 0,47 |
| | 85 | 0,69 | 0,98 | 0,81 | 0,31 | 0,84 | 0,45 | 0,83 | 0,42 | 0,56 |
| | 80 | 0,63 | 0,99 | 0,77 | 0,29 | 0,90 | 0,44 | 0,80 | 0,49 | 0,61 |

Tabla 5.3: Evaluación individual de las probabilidades usando el conjunto de datos ‘OrdonezA’.

| | | Conjunto de datos ‘OrdonezB’ | | | | | | | | |
|-----------|-----|------------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | | PAS | | | PSS | | | PDS | | |
| | BCI | PPV | VPR | FMeasure | PPV | VPR | FMeasure | PPV | VPR | FMeasure |
| Real | 98 | 0,95 | 0,41 | 0,57 | 0,90 | 0,25 | 0,39 | 0,99 | 0,79 | 0,86 |
| | 95 | 0,77 | 0,73 | 0,75 | 0,91 | 0,34 | 0,49 | 0,25 | 0,99 | 0,40 |
| | 90 | 0,45 | 0,91 | 0,60 | 0,87 | 0,40 | 0,55 | 0,11 | 0,99 | 0,20 |
| | 85 | 0,32 | 0,94 | 0,48 | 0,87 | 0,43 | 0,58 | 0,10 | 0,99 | 0,18 |
| | 80 | 0,25 | 0,94 | 0,39 | 0,82 | 0,45 | 0,58 | 0,08 | 0,99 | 0,15 |
| Sintético | 98 | 0,98 | 0,68 | 0,80 | 0,99 | 0,28 | 0,44 | 0,90 | 0,25 | 0,40 |
| | 95 | 0,93 | 0,86 | 0,90 | 0,90 | 0,43 | 0,59 | 0,90 | 0,34 | 0,49 |
| | 90 | 0,82 | 0,94 | 0,88 | 0,69 | 0,59 | 0,63 | 0,87 | 0,41 | 0,56 |
| | 85 | 0,69 | 0,98 | 0,81 | 0,31 | 0,84 | 0,45 | 0,83 | 0,42 | 0,56 |
| | 80 | 0,63 | 0,99 | 0,77 | 0,29 | 0,90 | 0,44 | 0,80 | 0,49 | 0,61 |

Tabla 5.4: Evaluación individual de las probabilidades usando el conjunto de datos ‘OrdonezB’.

| | | Conjunto de datos 'OrdonezC' | | | | | | | | |
|-----------|-----|------------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | | PAS | | | PSS | | | PDS | | |
| | BCI | PPV | VPR | FMeasure | PPV | VPR | FMeasure | PPV | VPR | FMeasure |
| Real | 98 | 0,91 | 0,52 | 0,66 | 0,24 | 0,88 | 0,38 | 0,79 | 0,50 | 0,62 |
| | 95 | 0,87 | 0,83 | 0,85 | 0,92 | 0,32 | 0,48 | 0,26 | 0,66 | 0,37 |
| | 90 | 0,75 | 0,95 | 0,84 | 0,84 | 0,49 | 0,62 | 0,25 | 0,82 | 0,39 |
| | 85 | 0,67 | 0,97 | 0,79 | 0,73 | 0,56 | 0,64 | 0,27 | 0,90 | 0,42 |
| | 80 | 0,60 | 0,98 | 0,75 | 0,68 | 0,62 | 0,64 | 0,30 | 0,98 | 0,45 |
| Sintético | 98 | 0,86 | 0,38 | 0,53 | 0,72 | 0,99 | 0,84 | 0,76 | 0,47 | 0,58 |
| | 95 | 0,75 | 0,87 | 0,80 | 0,21 | 0,99 | 0,35 | 0,25 | 0,64 | 0,36 |
| | 90 | 0,47 | 0,96 | 0,63 | 0,11 | 0,99 | 0,20 | 0,25 | 0,81 | 0,38 |
| | 85 | 0,32 | 0,98 | 0,48 | 0,06 | 0,99 | 0,12 | 0,26 | 0,89 | 0,40 |
| | 80 | 0,25 | 0,99 | 0,41 | 0,05 | 0,99 | 0,11 | 0,27 | 0,97 | 0,42 |

Tabla 5.5: Evaluación individual de las probabilidades usando el conjunto de datos 'OrdonezC'.

suelen estar balanceados, existiendo una desproporcion muy significativa en favor de los datos considerados no anómalos. En problemas binarios de estas características, donde los datos sufren de este tipo de desbalanceo, es muy común establecer criterios de valoración basados en rankings, como los que ofrece el área bajo la curva ROC (siglas de *Receiver Operating Characteristic*). Las curvas ROC son gráficos en los que se presenta los verdaderos positivos (sensibilidad) en función de los falsos positivos (complementario de la especificidad) para distintos puntos de corte. El parámetro para evaluar el rendimiento del modelo es el área bajo la curva (AUC), que tomará valores entre 1 (rendimiento perfecto) y 0,5 (rendimiento insignificante). El AUC es la medida considerada más adecuada para establecer comparativas entre sistemas de estas características [29].

Como se ha mencionado, el AUC está definida en términos de sensibilidad y especificidad. La sensibilidad nos indica la capacidad de nuestro estimador para detectar como casos positivos los patrones realmente atípicos. Es decir, la sensibilidad caracteriza la capacidad del modelo para detectar la anormalidad en un evento atípico. La especificidad nos indica la capacidad de nuestro estimador para detectar como casos negativos los patrones realmente normales. Es decir, la especificidad caracteriza la capacidad del modelo para detectar la ausencia de anomalías en patrones del comportamiento normales. Sensibilidad y especificidad se calculan como se detalla en las ecuaciones 5.27 y 5.28, respectivamente.

$$\text{Sensibilidad} = \frac{VP}{VP + FN} \quad (5.27)$$

$$\text{Especificidad} = \frac{VN}{VN + FP} \quad (5.28)$$

Puede apreciarse cómo la sensibilidad es equivalente a la Razón de Verdaderos Positivos. Aunque el porcentaje de aciertos no es una medida adecuada cuando se trabaja con datos desbalanceados, la incluimos también en las tablas, para obtener una comparativa más completa. El porcentaje de aciertos se define como se indica en la ecuación 5.29.

$$\text{Porcentaje de aciertos} = \frac{VP + VN}{VP + VN + FP + FN} \quad (5.29)$$

Haciendo uso de los parámetros que mejores resultados ofrecieron en el primer experimento se obtiene la salida del sistema. Las Figuras 5.3, 5.4 y 5.5 muestran el resultado de combinar las tres métricas como una única salida, para los conjuntos de datos ‘OrdonezA’, ‘OrdonezB’ y ‘OrdonezC’ respectivamente.

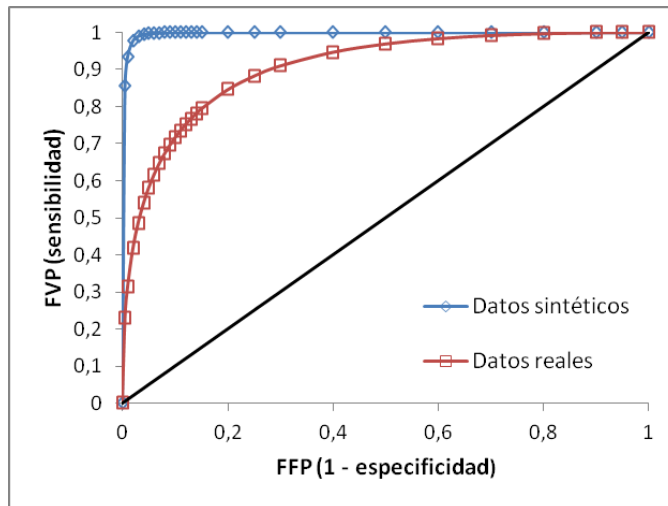


Figura 5.3: Curva ROC del modelo para el conjunto de datos ‘OrdonezA’.

El área bajo la curva muestra cómo el sistema ofrece un buen rendimiento para todos los conjuntos. Es especialmente ilustrativo la mejora en el rendimiento del modelo cuando se introducen errores sintéticos en los datos, independientemente del conjunto de datos empleado.

Para el conjunto de datos ‘OrdonezA’ se obtienen unos valores de AUC de 0,90 y 0,94 para los datos con errores reales y los datos con errores sintéticos, respectivamente. En el caso del conjunto ‘OrdonezB’ los valores del AUC son 0,90 y 0,98 para los datos con errores reales y sintéticos, respectivamente. Para el último conjunto, ‘OrdonezC’, nuestro modelo ofrece un rendimiento muy notable, obteniendo un AUC casi perfecto (0,99) para ambos tipos de errores.

Tal y como se aprecia en la Tabla 5.6, el modelo propuesto, con la configuración paramétrica obtenida del primer experimento, destaca especialmente identificando verdaderos negativos, es decir, detectando los comportamientos normales

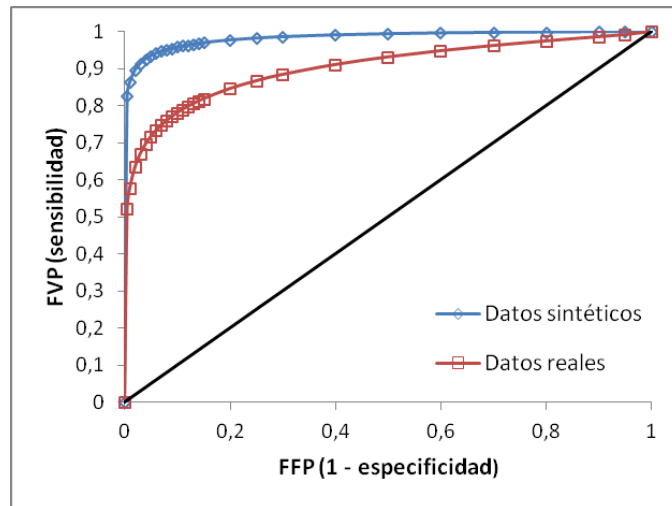


Figura 5.4: Curva ROC del modelo para el conjunto de datos 'OrdenezB'.

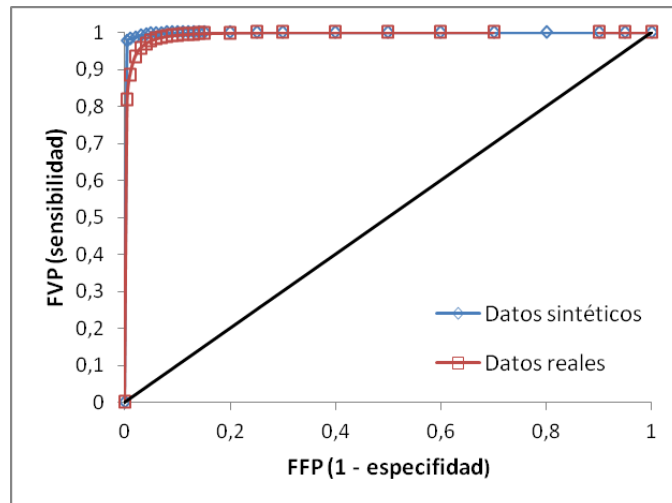


Figura 5.5: Curva ROC del modelo para el conjunto de datos 'OrdenezC'.

del usuario. Aunque en nuestro dominio puede resultar más interesante obtener una mayor sensibilidad en el modelo para prevenir situaciones anómalas. Respecto a esa premisa, debe tenerse en cuenta que, tal y como se ha mostrado en las tablas del apartado 5.2.3, al reducir el BCI el valor de la sensibilidad se incrementa notablemente, aunque ello suponga mermar el rendimiento global del sistema.

| | | Especificidad | Sensibilidad | Porcentaje de aciertos | AUC |
|-----------|----------|---------------|--------------|------------------------|------|
| Real | OrdonezA | 0,96 | 0,77 | 0,95 | 0,90 |
| | OrdonezB | 0,99 | 0,62 | 0,99 | 0,86 |
| | OrdonezC | 0,99 | 0,81 | 0,99 | 0,96 |
| | Promedio | 0,98 | 0,73 | 0,97 | 0,90 |
| Sintético | OrdonezA | 0,99 | 0,83 | 0,98 | 0,96 |
| | OrdonezB | 0,99 | 0,84 | 0,99 | 0,96 |
| | OrdonezC | 0,99 | 0,85 | 0,99 | 0,97 |
| | Promedio | 0,96 | 0,84 | 0,98 | 0,96 |

Tabla 5.6: Rendimiento para cada conjunto de datos, combinando las métricas.

5.2.5. Ejemplo de Funcionamiento

Para ilustrar cómo nuestra propuesta es capaz de diferenciar entre comportamientos anómalos y normales, en la Figura 5.6 se muestra un ejemplo de funcionamiento de la métrica PAS con el conjunto de datos ‘OrdonezB’. Se ha elegido dicho conjunto de datos debido a que ha sido generado por un usuario diagnosticado con la enfermedad de Parkinson, y por lo tanto, es el más interesante para evaluar la utilidad del sistema.

Cuando las personas diagnosticadas con la enfermedad de Parkinson llevan varios años de tratamiento, sus síntomas se generalizan y la medicación empieza a no ser del todo eficaz. Entonces, se comienza a producir un fenómeno llamado ‘on-off’ o ‘encendido-apagado’. Estos términos definen períodos de tiempo donde la sintomatología del paciente cambia bruscamente. En los períodos de ‘encendido’ el paciente se encuentra bien, ya que en ellos es cuando la medicación es efectiva. Pero cuando el efecto de la medicación disminuye sobrevienen bruscamente los síntomas de la enfermedad. A estos períodos de empeoramiento se les denomina períodos ‘off’ o ‘apagado’, y el paciente experimenta rigidez general y falta de coordinación en los movimientos. Sin embargo, durante los períodos ‘on’ el paciente se encuentra bien, se mueve fluidamente y está en control de sus movimientos. Las fluctuaciones entre los períodos ‘off’ y ‘on’ son impredecibles, y por tanto, pueden llegar a suponer un peligro real para el paciente.

El usuario que generó el conjunto ‘OrdonezB’ sufría de estos períodos ‘on-off’. En la Figura 5.6(b) se puede apreciar la probabilidad de activación para el

sensor de presión instalado en la cama. Estos valores se han obtenido empleando todos los días de datos de los que consta el conjunto. Desde las 11:00 hasta las 21:00 se puede apreciar cómo el usuario no hacía nunca uso de la cama, excepto durante 4 días en los cuales sufrió un periodo ‘off’. Se puede ver cómo el sistema identificó esos eventos como anomalías debido a la baja probabilidad de que el sensor se activase a esas horas.

Además, en la Figura 5.6(a) se muestra la probabilidad de activación del sensor magnético instalado en la puerta principal. Aparte de los eventos anómalos que se detectan en los límites definidos por el valor del BCI, en este caso es reseñable una anomalía muy evidente, que aparece durante una noche, entre las 03:00 y las 04:00. Esa activación del sensor se corresponde con un día en el que el usuario enfermó y tuvo que ser llevado al hospital. Por lo tanto, al abandonar la casa en un intervalo en el cual era muy poco probable que se activase el sensor de la puerta, el sistema identificó esta situación como un evento atípico.

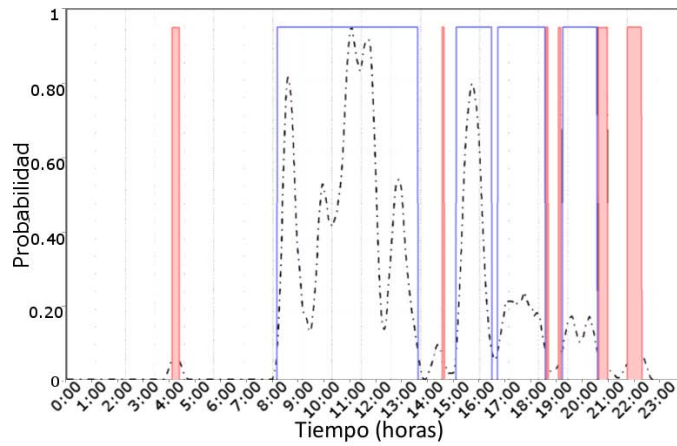
5.2.6. Comparativa

En la última parte del experimento se busca comparar el rendimiento de nuestra propuesta con otro sistema presente en la literatura, que ha demostrado ser útil en la detección de anomalías. Para la comparativa empleamos el modelo propuesto por Shin *et. al.*, que también emplea tres métricas para detectar los comportamientos anómalos de personas mayores que viven solas [216]. Estas métricas son calculadas empleando el nivel de actividad, el nivel de movilidad y los intervalos sin respuesta del usuario. Para identificar los comportamientos atípicos hace uso de una SVM, que recibe como entradas dichas métricas.

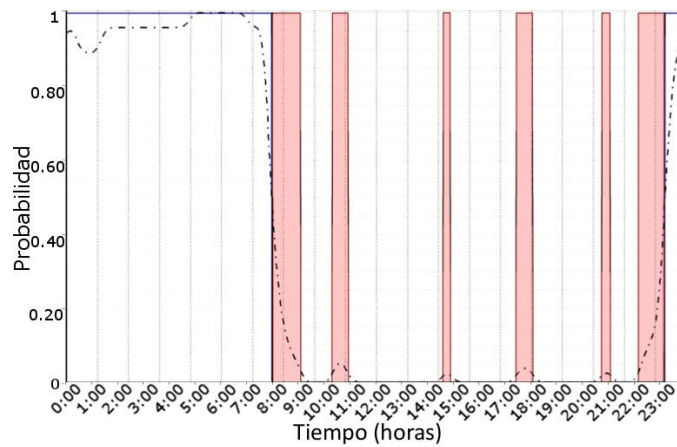
Al igual que en nuestro proceso de evaluación, Shin *et. al.* generan un conjunto de datos con errores sintéticos para cada conjunto de datos real. Aunque, los conjuntos de datos empleados para validar ambos modelos son diferentes, el enfoque es el mismo, y también emplean la especificidad, la sensibilidad y el área bajo la curva ROC como medidas de evaluación.

Por ello, se considera que dicha comparativa es adecuada y útil para determinar si los resultados obtenidos por el modelo propuesto son satisfactorios y comparables a otros sistemas de detección de anomalías presentes en la literatura. En la Tabla 5.7 se muestran los resultados para ambos modelos. La primera fila muestra el rendimiento de nuestra propuesta, donde las métricas PAS, PSS y PDS son empleadas para modelar el comportamiento. En la segunda fila se muestra el rendimiento del modelo presentado por Shin *et. al.*, que modelan el comportamiento mediante el nivel de actividad (AL), el nivel de movilidad (ML) y la longitud de los intervalos sin respuesta (NRI).

Para comprobar si las diferencias en rendimiento entre ambos modelos son significativas aplicamos como tests de significancia estadística la prueba *t* de Stu-



(a) PIR de la puerta principal



(b) Sensor de presión en la cama

Figura 5.6: Ejemplo de rendimiento para la métrica PAS en el conjunto de datos ‘Ordo-
nezB’ usando anomalías reales. Los espacios sombreados representan intervalos donde
comportamientos anómalos fueron detectados.

| Modelo | Datos reales | | Datos sintéticos | |
|-------------|---------------|--------------|------------------|--------------|
| Métricas | Especificidad | Sensibilidad | Especificidad | Sensibilidad |
| PAS&PSS&PDS | 0,98 | 0,73 | 0,96 | 0,84 |
| NRI&AL&ML | 0,85 | 0,74 | 0,85 | 0,86 |

Tabla 5.7: Comparativa entre nuestra propuesta y el sistema presentado por Shin *et. al.* [216].

dent ($p \leq 0,05$) para muestras pareadas y la prueba de los rangos con signo de Wilcoxon. La diferencia en sensibilidad entre ambos modelos no es estadísticamente significativa, para ninguna de las dos pruebas de significancia empleadas. Sin embargo, en especificidad nuestro modelo consigue un rendimiento significativamente superior al obtenido por Shin *et. al.* . Esta comparativa pone de manifiesto cómo nuestro modelo basado en análisis bayesiano obtiene resultados incluso mejores que otras propuestas ya establecidas. Para este caso en concreto, nuestro modelo tiene una capacidad similar de identificar eventos atípicos como tales, sin embargo, ofrece un rendimiento notablemente mejor en la identificación de eventos no anómalos.

5.3. Conclusiones

En este capítulo de la tesis doctoral hemos presentado y evaluado un método para la detección de anomalías en el comportamiento, que haciendo uso de tres métricas estadísticas, modela los patrones diarios de una persona y mediante análisis bayesiano realiza una estimación sobre qué eventos deben ser considerados anómalos.

El método está basado en una estimación aproximada de las métricas que definen los patrones diarios del usuario, y en unas probabilidades a priori que reflejan un conocimiento previo acerca de cómo deben ser dichas métricas. Éstas representan cuándo se activan los sensores de la red de monitorización, en qué orden y durante cuánto tiempo.

Se han demostrado tanto la eficacia como el valor del sistema mediante una evaluación en la que se han empleado tres conjuntos de datos obtenidos de tres entornos monitorizados diferentes. Las métricas presentadas han proporcionado información valiosa sobre los patrones del comportamiento de la cohorte monitorizada y el análisis bayesiano ha ofrecido un mecanismo muy consistente para manejar la incertidumbre del comportamiento humano.

La elección del intervalo de credibilidad bayesiana que gobierna la detección de anomalías es muy determinante en el rendimiento del modelo. En términos generales, la propuesta ofrece peor rendimiento en la detección de los verdaderos positivos (sensibilidad), sin embargo el sistema es suficientemente flexible para

orientar la detección hacia una mejora de la sensibilidad, en caso de requerirse así, aunque ello conlleve reducir el rendimiento global del modelo.

Al compararlo con otros métodos de detección de anomalías bien establecidos, se aprecia cómo nuestra propuesta ofrece mejores resultados, concretamente en términos de especificidad. Al igual que otros sistemas presentes en la literatura, nuestra propuesta considera los valores anómalos de los modelos como eventos extraños en el comportamiento de la persona, pero no llega a establecer una relación con el estado de salud del usuario, dejando esa tarea en manos de terceras personas, que sean las encargadas de determinar la relevancia médica de dichos eventos atípicos.

Al emplear estadística bayesiana, nuestro sistema permite integrar fácilmente el conocimiento previo que pudieramos tener sobre los parámetros que gobiernan las distribuciones de los eventos, permitiendo que el sistema se adapte a medida que se dispone de más datos generados por el usuario. El tiempo requerido por nuestra propuesta para modelar las métricas presentadas es muy pequeño, debido a que no hay que abordar ningún tipo de problema de convergencia en el aprendizaje, permitiendo además que los nuevos datos que se reciban desde los entornos instrumentados se puedan incorporar al modelo en tiempo real.

Los resultados sugieren que el sistema de detección de anomalías presentado, integrado dentro de nuestras redes de sensores, puede ser un valioso marco para llevar a cabo una monitorización de personas mayores que vivan de forma independiente. Además, gracias al uso de las probabilidades a priori, nuestro enfoque puede incorporar datos de diferentes fuentes de un modo sencillo, pudiendo ser potencialmente adecuado para ser empleado dentro de un esquema de transferencia de información.

Capítulo 6

Transferencia de Información

En el presente capítulo detallamos la solución propuesta en esta tesis doctoral para aplicar los métodos de modelado del comportamiento presentados hasta el momento en múltiples entornos monitorizados. En los capítulos previos, hemos presentado y detallado nuestras propuestas para modelar el comportamiento de una persona de un modo supervisado mediante reconocimiento de actividades y de un modo no supervisado mediante la detección de anomalías.

Hasta ahora, el enfoque propuesto es el de aplicar dichos modelos a un solo entorno, es decir, en una sola vivienda. Aunque los datos se han obtenido de diferentes viviendas, los modelos presentados se entrenan y evalúan de forma independiente sobre cada uno de los conjunto de datos. Teniendo en cuenta que la finalidad de estos modelos es la de ayudar a personas mayores a que vivan de forma independiente, se presupone que en el futuro este sistema pueda ser instalado en muchas y diferentes viviendas.

Sin embargo, debido a las diferencias tanto entre el diseño de casas como en el comportamiento de sus habitantes, un modelo optimizado para una vivienda no puede ser automáticamente empleado para otra vivienda diferente. El hecho de que un modelo obtenido en una vivienda no pueda ser transferido a otros entornos supone una fuerte limitación de estos sistemas.

El caso de los sistema de reconocimiento de actividades es un ejemplo ilustrativo de esta problemática, debido a que requieren de conjuntos de datos completamente etiquetados para llevar a cabo el proceso de inferencia. El aprendizaje supervisado plantea un problema cuando se quieren aplicar los clasificadores a mayor escala (por ejemplo, en múltiples viviendas). Debido a las diferencias en las distribuciones de las casas, y por tanto en los conjuntos de datos, un modelo generado para un entorno instrumentado concreto no puede ser usado en otro entorno diferente. Por esto, se necesita recoger suficientes datos en cada una de las viviendas donde se desee desplegar el sistema para poder generar un modelo personalizado. Este problema de escalabilidad limita la aplicabilidad de los métodos

de reconocimiento de actividades, ya que obtener nuevos datos etiquetados para cada nueva vivienda es costoso, y además no puede llevarse a cabo sin interferir en el comportamiento de los usuarios. En general, la obtención de una buena cantidad de datos etiquetados es difícil, puesto que los usuarios suelen ser reacios a escribir sus actividades, ya que requiere de bastante tiempo y de alguna manera compromete su privacidad. Esto es además aún más difícil para las personas con algún tipo de trastorno cognitivo, uno de los principales destinatarios de estos sistemas.

Dentro de la minería de datos, emplear un conocimiento adquirido durante un proceso de aprendizaje para mejorar el conocimiento previo acerca de procesos o problemas relacionados, es lo que se conoce como transferencia de información o *transfer learning*. Se diferencian una serie de tareas origen, de las cuales se extrae información, y una tarea objetivo, sobre la que se integran los datos obtenidos para mejorar el aprendizaje final de la misma. En nuestro dominio, la tarea objetivo se corresponde con la vivienda sobre la que se desea desplegar el sistema y de la que no disponemos de datos. La transferencia de información es relevante en nuestro dominio debido a que existen similitudes entre los diferentes modelos que se pueden obtener de distintas viviendas, y no es estrictamente necesario enfocar el modelado del comportamiento como una tarea específica del entorno.

En un enfoque probabilístico, es posible ver cada modelo entrenado en una vivienda como una instancia particular, y la distribución de estos modelos como la información a priori de nuestro sistema. Específicamente, en el reconocimiento de actividades se ha demostrado que es posible aprender la probabilidad a priori de los parámetros, usando las tareas secundarias, y emplear ese conocimiento para inferir los parámetros del modelo de la tarea principal [50], debido a que la probabilidad a priori permite obtener una buena estimación inicial de los parámetros que definirán el modelo de la tarea objetivo.

En la literatura existen ejemplos de cómo es posible integrar los valores a priori en el modelo de la tarea objetivo mediante el uso de un algoritmo de esperanza-maximización (EM) [130]. En una configuración de esperanza-maximización, un nuevo modelo puede ser inferido optimizando simultáneamente la probabilidad del modelo sobre la tarea objetivo y la probabilidad del modelo en relación a las distribuciones a priori. El problema de este enfoque, sin embargo, es que el mejor modelo obtenido para la tarea objetivo difícilmente es el modelo más probable según las dos condiciones indicadas. El resultado es que un modelo entrenado mediante un enfoque EM tiende a mostrar una *transferencia negativa*, donde la información obtenida de las tareas origen llega a reducir el rendimiento del modelo para la tarea objetivo. El fenómeno mediante el cual el uso de la transferencia de información tiene un efecto negativo en el proceso de aprendizaje de la tarea objetivo es conocido como transferencia negativa [37].

En la presente tesis doctoral postulamos que una solución eficiente a este pro-

blema es emplear estadística bayesiana para estimar los modelos y sus probabilidades a priori. Mediante el enfoque bayesiano, en vez de identificar el modelo de cada vivienda con una instancia particular, se considera la distribución sobre los posibles modelos a priori de las casas origen, y se usa esta información para inferir una distribución a priori sobre el modelo de la casa objetivo. Los modelos resultantes son más flexibles y tienen menos problemas de sobreadaptación, además de obtener un mejor rendimiento cuando se dispone de pocos datos de entrenamiento, tal y como demuestran nuestros experimentos.

En este capítulo presentamos un método de transferencia de información para nuestros problemas de reconocimiento de actividades y detección de anomalías que nos permite emplear el conocimiento adquirido en diferentes viviendas bajo un enfoque bayesiano, y sirve como base para la instalación de los sistemas en un entorno monitorizado completamente nuevo.

En el siguiente apartado detallamos la metodología empleada para llevar a cabo la transferencia de información con técnicas bayesianas. A continuación, evaluamos la propuesta tanto en el dominio del reconocimiento de actividades como en el de la detección de anomalías. Para cada parte describimos la configuración experimental y los resultados del modelo. Por último, finalizamos el capítulo con las conclusiones.

6.1. Metodología

Cuando se aplica la transferencia de información a los métodos de modelado del comportamiento, es común distinguir entre la vivienda o tarea objetivo, para la cual existen muy pocos o ningún dato de entrenamiento, y las viviendas o tareas origen, para las cuales sí que existen datos de entrenamiento. El problema a resolver es cómo se puede obtener información transferible desde un conjunto de casas, donde las redes de sensores sí han capturado datos útiles para entrenar nuestros modelos, para llevar a cabo el proceso de aprendizaje en la casa objetivo, donde la red de sensores todavía no ha comenzado a funcionar o no se dispone de datos etiquetados.

Aplicar la transferencia de información en la vivienda objetivo plantea fundamentalmente dos problemas:

- Cómo afrontar las diferencias entre las distintas redes de sensores derivadas de las diferentes configuraciones de las viviendas.
- Cómo inferir los parámetros de los modelos para que capturen información relativa al comportamiento de las personas.

6.1.1. Mapeado mediante Meta Características

El primer problema aborda el hecho de que es probable que existan diferencias en los espacios de características de los diferentes conjuntos de datos. Debido a que cada vivienda tiene una configuración diferente, cada red de sensores instalada también tendrá una topología diferente, dando lugar a su propio espacio de características. Para resolver esta cuestión se necesita introducir algún tipo de técnica de mapeado que permita establecer un espacio de características común para todas las tareas involucradas en el proceso de aprendizaje.

Tal y como se ha explicado en el Apartado 3.3, el espacio de características original de las viviendas está definido por el número de sensores instalados en las mismas, dado que cada sensor representa una característica. Por lo tanto, se necesita de una técnica de mapeado mediante la cual el espacio de características definido por los sensores, de dimensionalidad $N^{(i)}$, sea mapeado a un espacio de características común, de dimensión M .

Para poder transformar el espacio de características de los sensores mediante dicha técnica de mapeado hacemos uso de lo que en la literatura se denominan meta características [146], que son características que a su vez describen las propiedades de las auténticas características. Cada sensor está definido por una o más meta características, por ejemplo, un sensor instalado en el microondas puede ser representado por una meta característica que indique que el sensor está en la cocina, y por otra que indique que el sensor está instalado en un electrodoméstico para calentar alimentos.

Seleccionar una técnica de mapeado adecuada a nuestro problema es una tarea complicada, ya que una elección poco apropiada puede afectar notablemente al rendimiento del sistema. En anteriores estudios en relación a la transferencia de información se han llevado a cabo experimentos para comparar diversos tipos de mapeados [130]. La técnica de mapeado que obtuvo los mejores resultados fue la que combinada los sensores en base a su función (por ejemplo, los sensores empleados mientras se cocina). Para este trabajo se ha hecho uso del mismo tipo de mapeado, definiendo meta características que puedan describir la función de los sensores.

En nuestra propuesta, definimos un conjunto M de meta características y creamos manualmente un vector de mapeado $g_n^{(i)}$, el cual es un vector de valores binario que indica qué meta característica corresponde al sensor n de la vivienda i . Al juntar todos estos vectores de mapeado se obtiene una matriz de mapeado $G^{(i)} = \{g_n^{(i)}\}$, de dimension $N \times M$. En la Tabla 6.1 se detalla un ejemplo de cómo los vectores de los sensores (filas) son representados mediante meta características (columnas). Cada fila de valores binarios se corresponde al mapeado de un único sensor.

| Conjunto de datos | Sensor | Baño - otros | Baño - puerta | Cocina - calentar | Cocina - almacenamiento | Puerta entrada casa | Dormitorio - otros | Dormitorio - puerta | Inodoro |
|-------------------|-------------|--------------|---------------|-------------------|-------------------------|---------------------|--------------------|---------------------|---------|
| KasterenA | Microondas | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | Horno | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | ... | | | | | | | | |
| KasterenB | Microondas | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | Frigorífico | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| | ... | | | | | | | | |

Tabla 6.1: Ejemplo de sensores representados mediante el espacio de características original y el meta espacio.

6.1.2. Estimación de los Parámetros

Con el mapeado al meta espacio de características, se puede resolver el problema de las diferencias de dimensionalidad entre los espacios de características de los distintos conjuntos de datos. Con esta técnica podemos combinar los diferentes conjuntos de datos para estimar los parámetros $\theta^{(0)}$ de la vivienda objetivo.

Sin embargo, el segundo problema que plantea la transferencia de información es cómo tratar las diferencias en el comportamiento de los usuarios. Tanto en el problema de reconocimiento de actividades como en el de detección de anomalías, la distinción entre los datos objetivo y los datos origen es que los datos objetivo contienen información específica acerca de la vivienda sobre la que se quiere llevar a cabo el modelado del comportamiento, mientras que los datos origen contienen información más genérica. Un enfoque adecuado a este problema, que toma en cuenta expresamente los conocimientos previos que se pudieran tener sobre la tarea a resolver, es la estadística bayesiana.

En este problema, donde tenemos un modelo generado para cada una de las viviendas, y cada modelo con sus propios parámetros, el uso de las distribuciones a priori es una solución muy eficiente. Tal y como se ha mencionado anteriormente en este estudio, las distribuciones a priori cumplen un papel importante en el análisis bayesiano ya que miden el grado de conocimiento inicial que se tiene de los parámetros en estudio. Si bien su influencia disminuye a medida que más información muestral está disponible, el uso de una u otra distribución a priori determina ciertas diferencias en la distribución a posteriori.

De cada una de las viviendas origen se obtiene una distribución a priori de los parámetros del modelo, ofreciendo una estimación inicial razonable de los valores de los parámetros para la vivienda objetivo. A continuación, los datos obtenidos en la vivienda objetivo se van añadiendo progresivamente a las distribuciones a

priori del modelo, disminuyendo la influencia de los datos origen.

Un diagrama general de la propuesta puede apreciarse en la Figura 6.1, donde θ representa los parámetros del modelo de reconocimiento de actividad o de detección de anomalías y Ψ representa los parámetros de las distribuciones a priori, los cuales son denominados como hiperparámetros. Para cada casa i , los datos de los sensores se emplean para aprender los parámetros $\theta^{(i)}$ de modelo. Todos los modelos origen se emplean para estimar los hiperparámetros Ψ de las distribuciones a priori, que posteriormente se usan para aprender los parámetros $\theta^{(R)}$ del modelo objetivo.

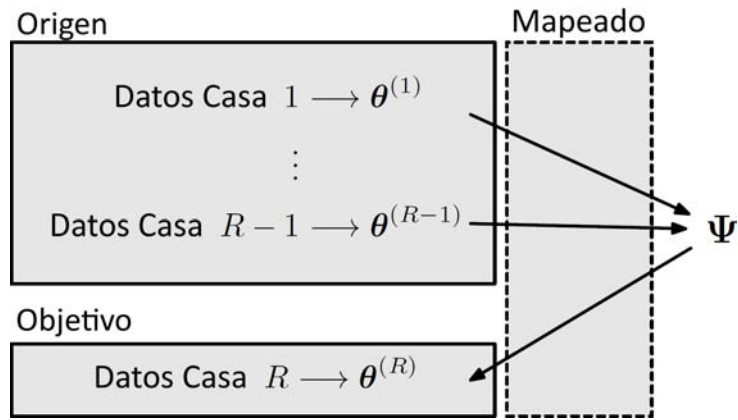


Figura 6.1: Esquema de la propuesta para realizar la transferencia de información.

El modo de integrar las distribuciones a priori en el proceso de aprendizaje dependerá, lógicamente, del problema que estemos abordando.

- Para llevar a cabo el reconocimiento de actividades empleamos un modelo oculto de Márkov entrenado mediante un proceso de inferencia bayesiana, tal y como se describe en el Apartado 4.4 de esta tesis doctoral. Mediante este mecanismo es posible incluir los datos de las viviendas origen como valores a priori en el algoritmo Montecarlo basado en cadenas de Markov (MCMC).
- Para el problema de detección de anomalías empleamos el método basado en probabilidades que se presentó en el Apartado 5.1. Dicho método, al hacer uso de la estadística bayesiana para la estimación de las métricas, es capaz de integrar los datos de las viviendas origen como parte de sus valores a priori.

En cada uno de los problemas los parámetros que gobiernan el modelo son diferentes (ver Tablas 4.11 y 5.1), pero sin embargo comparten el método para

aplicar la transferencia de conocimiento. En ambos casos el primer paso es estimar los parámetros de los modelos para las casas origen, a continuación se decide qué distribuciones de probabilidad se emplearán para representar las distribuciones a priori y se usan los parámetros para el cálculo de los hiperparámetros, y finalmente los valores a priori se introducen en el proceso de aprendizaje empleando el marco bayesiano.

Los valores de los hiperparámetros pueden ser fácilmente interpretables si estos se eligen adecuadamente o, como es nuestro caso, se emplean las distribuciones conjugadas como distribuciones a priori. Por ejemplo, en el caso del reconocimiento de actividades, los valores de la distribución beta que define la probabilidad a priori de la observación de un sensor pueden ser interpretados como el número de observaciones de ese sensor. Del mismo modo, los valores de los parámetros de la distribución de Dirichlet pueden ser interpretados como el número efectivo de transiciones entre dos sensores.

Debido a que los parámetros de las distribuciones a priori son estimados a partir de todas las viviendas origen, la operación que se lleva a cabo es básicamente es un sumatorio de todas las esperanzas de los valores. Cómo de relevante acaba siendo un valor a priori obtenido de un conjunto de datos origen en el conjunto de datos casa objetivo dependerá del nivel de similitud de la esperanza para ese valor en concreto.

Por ejemplo, si en todas las casas el microondas se ha usado cuando se cocinaba, entonces la distribución a priori estará sesgada hacia la derecha, con los valores muy concentrados cerca del 1, representando que es altamente probable que el microondas se active mientras se cocina. Esto dará lugar a un valor alto para el hiperparámetro (Figura 6.2(a)), e influirá notablemente en la esperanza de la vivienda objetivo. Por otro lado, si los datos de las casas objetivo no son similares, eso dará lugar a una distribución a priori muy aplanada (Figura 6.2(b)) con un valor bajo para el hiperparámetro, y por lo tanto influirá poco en la esperanza de la vivienda objetivo.

6.2. Evaluación

En este apartado se detalla más la propuesta de transferencia de información tanto para el reconocimiento de actividad como para la detección de anomalías y se presentan los resultados de la evaluación.

Para el problema de reconocimiento de actividades, se evalúa la propuesta sobre los cinco conjuntos de datos totalmente etiquetados que se han empleado previamente en esta tesis doctoral. Además de presentar los resultados de aplicar la transferencia de información a nuestros datos, se examina la pérdida de precisión que supone transformar el espacio de características mediante la técnica

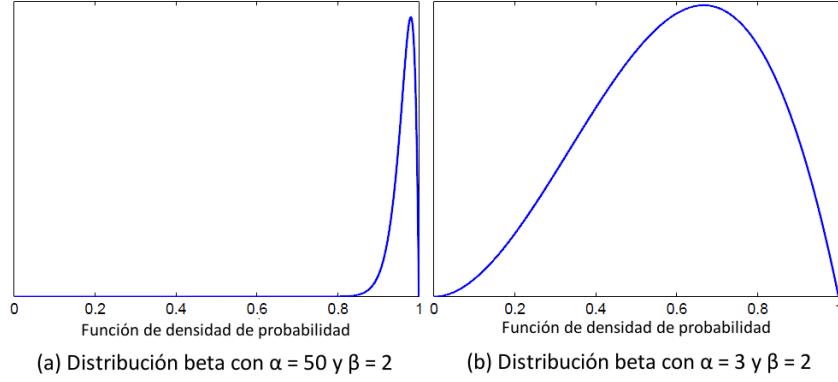


Figura 6.2: Ejemplos de una distribución beta puntiaguda (a) y aplanada (b).

de mapeado propuesta. Por último, también se presenta una comparativa con otra técnica de transferencia de información propuesta para el mismo problema, que igualmente hace uso de las probabilidades a priori, pero cuyo proceso de aprendizaje se basa en el algoritmo EM.

Para el problema de detección de anomalías, se evalúa el método sobre los tres conjuntos de datos empleados en el capítulo anterior. Al igual que para el reconocimiento de actividades, también se lleva a cabo una evaluación para examinar qué pérdida de precisión supone transformar el espacio de características.

6.2.1. Reconocimiento de Actividades

Tal y como se ha indicado previamente, el modelo de reconocimiento de actividad sobre el que se aplica la transferencia de información en este trabajo es el HMM entrenado mediante inferencia bayesiana a través del algoritmo MCMC.

Para llevar a cabo el reconocimiento de actividad se emplea el mismo HMM que se presentó en el Capítulo 4. A continuación se incluye un breve resumen de los detalles de dicho HMM. La probabilidad conjunta del modelo se define mediante la ecuación 6.1.

$$p(\mathbf{y}_{1:T}, \mathbf{x}_{1:T}) = p(\mathbf{y}_1) \prod_{t=1}^T p(\vec{\mathbf{x}}_t | \mathbf{y}_t) \prod_{t=2}^T p(\mathbf{y}_t | \mathbf{y}_{t-1}). \quad (6.1)$$

En donde cada uno de los diferentes factores representan: la distribución de estados iniciales $p(\mathbf{y}_1)$, representada mediante una distribución multinomial y parametrizada por π ; la distribución de observación $p(\vec{\mathbf{x}}_t | \mathbf{y}_t)$ que se compone de un conjunto de distribuciones de Bernoulli independientes, parametrizadas por $\mathbf{B} = \{\mu_{in}\}$; la distribución de transición $p(\mathbf{y}_t | \mathbf{y}_{t-1})$, representada como una colección de distribuciones multinomiales, parametrizadas por $\mathbf{A} = \{a_i\}$. Por

lo tanto, el modelo completo es parametrizado por el conjunto de parámetros $\theta = \{\pi, \mathbf{A}, \mathbf{B}\}$.

Tal y como se recoge en la Tabla 4.11, las distribuciones a priori para los parámetros de la distribución de estados iniciales y la distribución de transición se asumen que son distribuciones de Dirichlet (con hiperparámetros η y ρ). Mientras que la distribución a priori para la distribución de observaciones es una beta (con hiperparámetros ω, ν). Por lo tanto, el conjunto de hiperparámetros para nuestro modelo de reconocimiento de actividad está definido por $\Psi = \{\eta, \rho, \omega, \nu\}$.

Los valores de los hiperparámetros son estimados dentro del algoritmo MCMC detallado en el Apartado 4.4.1. Los hiperparámetros η y ρ se estiman tal y como se indica en la ecuación 4.25. Los hiperparámetros ω, ν se calculan según las ecuaciones 4.26 y 4.27.

Los hiperparámetros estimados empleando las viviendas origen son los que gobiernan totalmente las distribuciones a priori nada más empezar el algoritmo MCMC. Durante las primeras iteraciones del algoritmo, los primeros pasos del mismo (las actualizaciones de los parámetros) estarán totalmente condicionados por dichos valores a priori. Sin embargo, a medida que se introduzcan más datos procedentes de la vivienda objetivo, la influencia de esos valores irá disminuyendo.

Se debe destacar que la lista de AVD debe ser la misma para todas las viviendas que se incluyen en el proceso de transferencia de información, debido a que para las etiquetas no existe ningún tipo de mapeo y el espacio debe ser el mismo. Es por ello que para un conjunto de datos objetivo sólo se pueden usar los conjuntos de datos origen que compartan las mismas clases.

Conjuntos de Datos

La experimentación se ha realizado empleando los cinco conjuntos de datos descritos en el Capítulo 4. En la Tabla 6.2 se detallan dichos conjuntos, indicando el número de instancias que los componen y cuántas actividades han sido etiquetadas para cada uno. Conocer el número de instancias que compone cada conjunto de datos es ilustrativo para analizar el peso que puede llegar a tener dicho conjunto en el proceso de aprendizaje.

Tal y como se ha mencionado anteriormente, en nuestro planteamiento el número de actividades para cada uno de los conjuntos de datos que toman parte en la transferencia de información debe ser el mismo. Es por ello, que debido a diferencias en el espacio de actividades, no todos los conjuntos puede servir como conjunto de datos origen. Específicamente, en nuestra experimentación, usamos los conjuntos de datos ‘KasterenA’, ‘KasterenB’ y ‘KasterenC’ en un experimento y los conjuntos de datos ‘OrdenezA’ y ‘OrdenezC’ en otro diferente. Es decir, cuando evaluamos el método sobre el conjunto ‘KasterenA’ se emplearán los con-

juntos ‘KasterenB’ y ‘KasterenC’ para obtener los valores a priori, pero no los conjuntos ‘OrdenezA’ y ‘OrdenezC’, debido a sus diferentes espacios de actividades.

Tabla 6.2: Conjuntos de datos empleados en la transferencia de información.

| Conjunto de datos | KasterenA | KasterenB | KasterenC | OrdenezA | OrdenezC |
|-----------------------|-----------|-----------|-----------|----------|----------|
| Número de instancias | 35486 | 19968 | 26236 | 20160 | 30240 |
| Sensores | 14 | 22 | 21 | 12 | 12 |
| Número de actividades | 8 | 8 | 8 | 11 | 11 |

Configuración de la Experimentación

Como se ha indicado, un HMM es el modelo empleado para llevar a cabo el reconocimiento de actividad en todos los experimentos. La técnica de mapeado empleada para obtener un espacio de características común se realiza según la función de los sensores. El mecanismo de mapeado se describe más detalladamente en el Apéndice B.

El flujo de datos de los sensores se discretiza en intervalos de longitud $\Delta t = 60$ segundos. Esta granularidad es suficientemente extensa para ser discriminativa y suficientemente breve para mantener una buena resolución de etiquetado. Tras la discretización se obtiene un número de instancias para cada conjunto de datos tal y como viene detallado en la Tabla 6.2.

Los conjuntos de datos son divididos en entrenamiento y test mediante un tipo de validación cruzada ‘dejando uno fuera’, y el resultado final es un promedio de todas las iteraciones.

Las métricas para evaluar el método son el Valor Predictivo Positivo (PPV), la Razón de Verdaderos Positivos (VPR) y la medida F o *F-Measure*. Estas métricas se detallaron previamente en el Apartado 4.3.2, y resultan especialmente útiles cuando se debe tratar con conjuntos de datos no balanceados, como es el caso.

El tipo de representación de datos empleada es la representación de punto de cambio o *changepoint* (ver Apartado 4.3.2). Se ha elegido esta representación con el objetivo de poder comparar nuestro modelo con la propuesta presentada en [130], donde se presentó un sistema similar, pero que empleaba una variación del algoritmo esperanza-maximización para integrar los valores de hiperparámetros en el proceso de aprendizaje.

Para estimar si las diferencias en precisión son estadísticamente significativas empleamos un test estadístico *t* de Student con nivel de significación de 0.05 y la prueba de los rangos con signo de Wilcoxon.

Resultados

El método de transferencia de información propuesto nos permite afrontar un inconveniente que plantean los sistemas de reconocimiento de actividades, que es la necesidad de datos debidamente etiquetados para cada entorno donde se quiera desplegar el sistema. En este apartado, presentamos los experimentos realizados para evaluar la propuesta empleando conjuntos de datos reales.

En esta experimentación primero evaluamos cómo afecta la transformación del espacio de características al rendimiento del sistema. Para este primer experimento no se emplea ningún tipo de transferencia de información, sino que se compara el rendimiento de la propuesta usando el espacio de características original (espacio de los sensores) y el transformado mediante la función de mapeo (meta espacio). Es decir, cada conjunto de datos se evalúa tomando datos únicamente de la vivienda donde se generó. El mapeado se realiza tal y como se ha descrito anteriormente y el aprendizaje de los parámetros del HMM se lleva a cabo usando la inferencia bayesiana mediante MCMC que se ha detallado en el Apartado 4.4.

Los resultados para los cinco conjuntos de datos se muestran en la Tabla 6.3. Como se puede apreciar el rendimiento del sistema cuando se usa el meta espacio de datos es menor para todos los conjuntos, debido a la pérdida de información que supone realizar el mapeado. Aunque según los tests de significancia la mayoría de las diferencias no llegan a ser estadísticamente significativas, sí que se observa una pérdida consistente en los valores de las métricas.

Tabla 6.3: Rendimiento del HMM entrenado mediante MCMC usando el espacio de características original y el mapeado.

| Conjunto de datos | Espacio de características | PPV | VPR | FMeasure |
|-------------------|----------------------------|------|------|----------|
| KasterenA | Sensores | 0,81 | 0,70 | 0,75 |
| | Meta | 0,79 | 0,65 | 0,71 |
| KasterenB | Sensores | 0,78 | 0,78 | 0,78 |
| | Meta | 0,75 | 0,71 | 0,72 |
| KasterenC | Sensores | 0,67 | 0,58 | 0,62 |
| | Meta | 0,60 | 0,59 | 0,59 |
| OrdonezA | Sensores | 0,79 | 0,62 | 0,69 |
| | Meta | 0,65 | 0,61 | 0,63 |
| OrdonezC | Sensores | 0,75 | 0,71 | 0,72 |
| | Meta | 0,72 | 0,67 | 0,69 |

Este resultado no puede considerarse inesperado ya que el espacio de caracte-

rísticas de los sensores es específico de la vivienda, mientras que el meta espacio requiere de una generalización de las características que implica pérdida de información. No obstante, debido a la necesidad de establecer un espacio común para aplicar la transferencia de información, el empleo de las meta características sirve como una solución válida para tratar con las diferentes configuraciones de las redes de sensores implicadas, en el sentido que permiten un rendimiento competitivo cuando se comparan con el espacio de características original.

El segundo experimento relativo al uso de transferencia de información en el reconocimiento de actividades muestra el rendimiento del modelo para los cinco conjuntos de datos empleados y establece una comparativa con otra técnica similar cuyo método de aprendizaje está basado en el algoritmo EM. En este experimento una de las viviendas se emplea como conjunto de datos objetivo y el resto sirve como conjuntos origen, siempre y cuando compartan el espacio de actividades, tal y como se ha indicado anteriormente. Por ello, los resultados se muestran mediante dos pruebas diferentes, en la primera toman parte los conjuntos de datos ‘KasterenA’, ‘KasterenB’ y ‘KasterenC’, mientras que para la segunda se usan los conjuntos ‘OrdonezA’ y ‘OrdonezC’.

Además de nuestra propuesta, en la comparativa se muestran los resultados de una técnica de transferencia de información basada en el algoritmo EM, que ha demostrado buenos resultados en este dominio [130]. En dicha propuesta, el algoritmo EM se emplea para estimar el modelo más probable o MAP (*maximum a posteriori*). Durante el primer paso del método, todos los datos disponibles en los conjuntos de entrenamiento son empleados para calcular la esperanza mediante el algoritmo *forward-backward* [197]. En el segundo paso del método, o etapa de maximización, los parámetros del modelo se estiman mediante una serie de ecuaciones que incorporan las probabilidades a priori [110]. Este algoritmo se ejecuta hasta alcanzar la convergencia.

Esta estimación para obtener los parámetros MAP mediante EM demostró ser apropiada para llevar a cabo el aprendizaje de los HMM en el dominio de la transferencia de información, especialmente cuando se disponía de pocos o ningún dato para la vivienda objetivo. Sin embargo, esta propuesta comparte los inconvenientes propios del algoritmo EM: tiene un rendimiento muy condicionado a los valores iniciales de los parámetros, tiende a converger a máximos locales y el algoritmo *forward-backward* es numéricamente sensible.

En la comparativa también se introducen los valores obtenidos por el modelo cuando no se realiza ningún tipo de transferencia de información. De este modo se puede analizar qué método de transferencia de información ofrece los mejores resultados y cómo varía el rendimiento al emplear la transferencia de información.

En la Figura 6.3 se muestran los resultados para el experimento con los conjuntos ‘KasterenA’, ‘KasterenB’ y ‘KasterenC’, y en la Figura 6.4 los resultados para el experimento con los conjuntos ‘OrdonezA’ y ‘OrdonezC’. El eje X muestra

el número de días de datos que se emplean en el proceso de aprendizaje. La observación más inmediata es que ambos métodos de transferencia de información son capaces de llevar a cabo el reconocimiento de actividades cuando no se emplea ningún dato de la vivienda objetivo en el proceso de aprendizaje (el número de días es igual a cero). Lógicamente, el método específico a la vivienda no dispone de datos etiquetados con los que estimar los parámetros, mientras que los métodos de transferencia de información pueden utilizar los datos de los conjuntos origen.

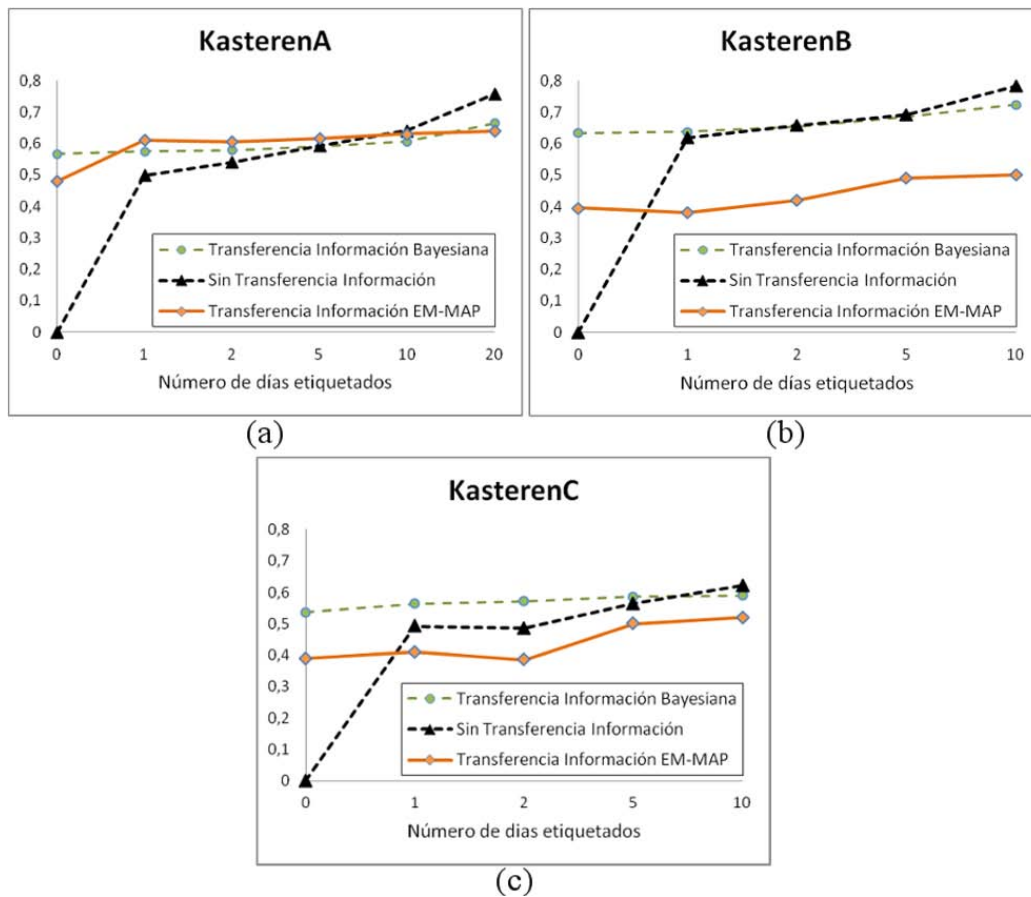


Figura 6.3: Primer experimento comparando los modelos EM-MAP y MCMC para realizar la transferencia de información en el reconocimiento de actividades.

Incluso cuando se dispone de pocos datos para la vivienda objetivo, nuestra propuesta bayesiana ofrece un mejor rendimiento que el método sin transferencia de información. Dicha diferencia de rendimiento se reduce a medida que más datos de la vivienda objetivo se introducen en el proceso de aprendizaje, mostrando cómo el uso de los hiperparámetros es útil para determinar los parámetros del HMM cuando se dispone de pocos datos. Cuantos más datos de la vivienda obje-

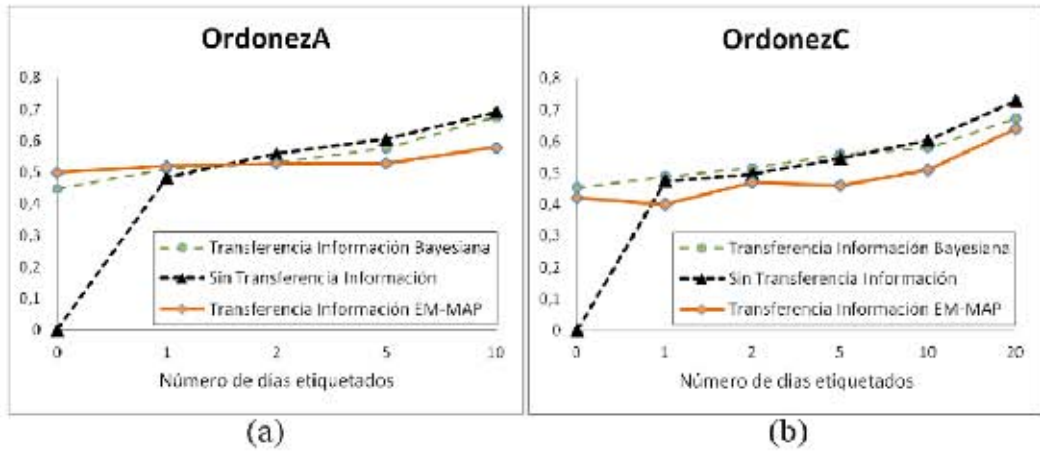


Figura 6.4: Segundo experimento comparando los modelos EM-MAP y MCMC para realizar la transferencia de información en el reconocimiento de actividades.

tivo se emplean para la estimación de los parámetros, menor influencia tienen los valores a priori, y los datos locales pasan a dominar el aprendizaje.

Las pruebas de significancia muestran cómo las diferencias de rendimiento entre los dos métodos de transferencia de información son estadísticamente significativas para los conjuntos ‘KasterenB’, ‘KasterenC’ y ‘OrdenezC’, ofreciendo la propuesta bayesiana mejores resultados. En los conjuntos ‘KasterenA’ y ‘OrdenezA’ las diferencias no son significativas. No obstante, la propuesta bayesiana obtiene mejores resultados que el método EM-MAP para todos los conjuntos de datos. Esta mejora se produce porque el aprendizaje bayesiano no tiende a la sobreadaptación [8], incluso cuando se dispone de pocos datos de entrenamiento. Es más, los algoritmos de máxima verosimilitud y esperanza-maximización pueden converger hacia máximos locales, obteniendo resultados muy pobres. Al emplear una técnica de muestreo en vez de una de maximización, se evita dicho riesgo.

El hecho de que la propuesta bayesiana no obtenga una mejora significativa respecto a la técnica EM-MAP para los conjuntos de datos ‘KasterenA’ y ‘OrdenezA’, se debe muy probablemente a que dichos conjuntos tienen el etiquetado más metódico, puesto que ambos fueron generados por investigadores. Posiblemente esa sea la causa de que el método EM-MAP mejore su rendimiento hasta casi obtener los valores de la transferencia bayesiana. Esto implicaría que el método bayesiano resulta más adecuado para tratar con un etiquetado de peor calidad, demostrando que la técnica propuesta es más robusta frente a los datos ruidosos.

6.2.2. Detección de Anomalías

El método de transferencia de información para detección de anomalías emplea la misma metodología que el sistema presentado en el Capítulo 5. Este método aplica un enfoque bayesiano para, en función de tres métricas, obtener una estimación probabilística del comportamiento de los usuarios.

Para llevar a cabo la detección de anomalías se hace uso de tres tipos probabilidades, que se obtienen de la información ofrecida por la red de sensores, a saber: probabilidad de activación, probabilidad de secuencia y probabilidad de duración de los sensores.

Cada una de estas probabilidades se define de acuerdo a una distribución diferente, tal y como se recoge en la Tabla 5.1. La probabilidad de activación de un sensor se define mediante una distribución de Bernoulli parametrizada por ρ , la probabilidad de secuencia se representa mediante una distribución multinomial parametrizada por θ y la probabilidad de duración de los sensores se representa mediante una distribución normal parametrizada por μ y λ . Por lo tanto, el modelo de detección de anomalías es parametrizado de forma completa por el conjunto de parámetros $\theta = \{\rho, \theta, \mu, \lambda\}$.

Las distribuciones a priori para los parámetros de cada una de dichas distribuciones de probabilidad, son aquellas que pertenecen a la familia conjugada. La distribución a priori de los parámetros de la probabilidad de activación es una distribución Beta, parametrizada por α y β . Los valores a priori de los parámetros de la distribución multinomial de la probabilidad de secuencia se definen mediante una distribución de Dirichlet, parametrizada por η . Finalmente, la distribución a priori para la probabilidad de duración es una distribución Normal-Gamma, con parámetros ϕ, κ, a y b . Esto es, el conjunto de hiperparámetros del modelo es el compuesto por $\Psi = \{\alpha, \beta, \eta, \phi, \kappa, a, b\}$.

Los valores de los hiperparámetros son estimados tal y como se recoge en el Apartado 5.1. El cálculo para la probabilidad de activación se indica mediante las ecuaciones 5.4 y 5.5, para la probabilidad de secuencia mediante la ecuación 5.10, y para la probabilidad de duración mediante las ecuaciones 5.18, 5.19, 5.20 y 5.21. Debe tenerse en cuenta que, considerando estos cálculos, ahora los valores iniciales para los parámetros de la distribución a priori de la tarea objetivo no son difusos, sino que vienen determinados para las tareas origen.

Al igual que pasa con el problema de reconocimiento de actividades, cuando se dispone de pocos datos en la vivienda objetivo para entrenar el modelo, los parámetros estarán totalmente condicionados por los valores a priori. No obstante, a medida que se introduzcan más datos en el proceso de estimación, la influencia de esos valores deberá disminuir.

Conjuntos de Datos

La evaluación de este método se ha realizado empleando los tres conjuntos de datos descritos en el Apartado 5.2.1. Estos conjuntos son los descritos en la Tabla 3.1.

Para esta experimentación se han empleado las versiones de los conjuntos que únicamente contienen anomalías reales. Dichos eventos atípicos son los mismos que se emplearon para evaluar el método de detección de anomalías, por lo que también han sido identificados de forma manual por los investigadores.

Configuración de la Experimentación

La división de los datos en entrenamiento y test se ha realizado mediante una validación cruzada del tipo ‘dejando uno fuera’. El número de iteraciones de esta validación depende del número de días que defina al conjunto de datos objetivos. En cada iteración se hace uso de una submuestra de datos diferente y, al final, se obtiene un resultado promedio. Para esta experimentación la técnica de mapeado empleada para obtener un espacio de características común se realiza también según la función de los sensores. El mecanismo de mapeado se describe más detalladamente en el Apéndice B.

Tal y como se ha explicado previamente, para poder estimar las métricas PAS y PDS el flujo de datos tiene que estar discretizado en un conjunto de intervalos. Esta discretización se ha llevado a cabo empleando una granularidad $\Delta t = 60$ segundos.

Para evaluar este método también se emplea el Valor Predictivo Positivo (PPV), la Razón de Verdaderos Positivos (VPR) y la medida F o *F-Measure*. Estas métricas se calculan tal y como se indicó en el Apartado 5.2.3. Estas medidas son adecuadas para el problema debido al inherente desbalanceo que sufren los conjuntos de datos destinados a la detección de anomalías.

Al igual que para el reconocimiento de actividades, para estimar si las diferencias en precisión son estadísticamente significativas empleamos un test estadístico *t* de Student con nivel de significación de 0.05 y la prueba de los rangos con signo de Wilcoxon.

Resultados

Aplicar la transferencia de información a la detección de anomalías nos permite emplear el método en una vivienda desde el mismo momento que se despliega el sistema, es decir, cuando todavía hay muy pocos datos procedentes de la vivienda objetivo. En este apartado, presentamos los experimentos llevados a cabo para evaluar el método de transferencia de información empleando los conjuntos de datos reales.

La experimentación es similar a la realizada para evaluar la propuesta de reconocimiento de actividad. El primer experimento está enfocado a determinar cómo afecta la transformación del espacio de características al rendimiento del modelo. Este experimento es parecido al presentado en el Apartado 5.2.4, donde no se aplica ningún tipo de transferencia de información, aunque en este caso se compara el rendimiento de la propuesta usando el espacio de características original (espacio de los sensores) y el transformado mediante la función de mapeo (meta espacio). La transformación de los datos al meta espacio se realiza tal y como se ha indicado previamente.

Los resultados para este experimento se muestran en la Tabla 6.4. Al igual que en el caso del reconocimiento de actividad, el rendimiento del meta espacio de datos es inferior al del espacio de datos original. Sin embargo, en este caso las diferencias son más notables y estadísticamente significativas, según la prueba de Student y la prueba de Wilcoxon.

Tabla 6.4: Rendimiento del modelo de detección de anomalías usando el espacio de características original y el mapeado.

| Conjunto de datos | Espacio de características | PPV | VPR | FMeasure |
|-------------------|----------------------------|------|------|----------|
| OrdenezA | Sensores | 0,64 | 0,87 | 0,73 |
| | Meta | 0,62 | 0,44 | 0,51 |
| OrdenezB | Sensores | 0,86 | 0,65 | 0,73 |
| | Meta | 0,73 | 0,46 | 0,56 |
| OrdenezC | Sensores | 0,78 | 0,66 | 0,71 |
| | Meta | 0,73 | 0,40 | 0,50 |

Esta notable diferencia se debe a que el método de detección de anomalías está basado principalmente en estimar directamente el espacio de características mediante diferentes probabilidades. Por esto, si se modifica dicho espacio se debería volver a analizar los datos y determinar qué eventos deben ser considerados anómalos. Para que el sistema pudiera obtener un rendimiento con las meta características similar al conseguido con el espacio de datos original, todos los sensores que compartan una misma función deberían modelarse igual. Si por ejemplo hay dos sensores que en el meta espacio representan la misma característica (porque los dos se relacionan con la acción de cocinar, por ejemplo), pero la forma que el usuario tiene de interactuar con ellos es diferente (porque uno lo active al preparar el desayuno y otro al preparar la cena, por ejemplo), la transformación de los datos mediante la función de mapeo implica la pérdida de esa información. Como resultado de esto, el sistema podría detectar parte de los eventos del meta espacio

como falsos negativos, como efectivamente sucede, tal y como indica el hecho de que la mayor reducción en rendimiento sea por parte de la VPR.

A la vista de los resultados se aprecia cómo nuestro método de detección de anomalías es muy sensible a la transformación del espacio de características, y emplear una función de mapeo basada en la función de los sensores posiblemente no sea la mejor opción en este caso. No obstante, aunque el tipo de mapeado elegido implica una notable pérdida de información, nos sirve para comprobar si la técnica de detección de anomalías se adapta a la transferencia de información.

En el segundo experimento relativo al uso de transferencia de información en la detección de anomalías, se evalúa el rendimiento del método para los tres conjuntos de datos empleados. Para este experimento cada una de las viviendas se emplea como vivienda objetivo una vez y el resto sirve como viviendas origen. En la Figura 6.5 se muestran los diferentes valores obtenidos por el modelo. El eje X muestra el número de días de datos de la vivienda objetivo que se incluyen en el aprendizaje.

Se puede apreciar cómo la transferencia de información es especialmente útil cuando se dispone de muy pocos datos de la vivienda objetivo, ya que permite comenzar el proceso de aprendizaje empleando los valores obtenidos de las otras viviendas.

A medida que se emplean más datos procedentes de la vivienda objetivo, la mejora ofrecida por la transferencia de información se reduce. De hecho, cuando se dispone de pocos datos del conjunto objetivo, la mejora ofrecida por el método de transferencia no es significativa. A medida que se dispone de más datos de la vivienda objetivo, menor influencia deben tener los valores a priori. Sin embargo, en este caso se observa cómo el modelo tiende a mostrar una transferencia negativa cuando se dispone de más datos de la vivienda objetivo. Los valores a priori condicionan el aprendizaje de tal modo que llegan a reducir el rendimiento del modelo para la tarea objetivo.

Al basar la detección de las anomalías en una estimación estadística de las características, los valores a priori cobran más importancia que en otros enfoques. Aunque es altamente probable que la influencia de dichos valores a priori se vaya diluyendo cuando los datos del conjunto objetivo aumenten, ese hecho no puede confirmarse con nuestro experimento. Por lo que se puede concluir que la transferencia de información en este dominio sí es útil, pero sólo cuando se dispone de muy pocos datos objetivos, ya que posteriormente se da lugar a una influencia negativa, tal y como está definido el método de transferencia.

No obstante, un interesante trabajo futuro sería aplicar algún tipo de mecanismo que pudiera reducir paulatinamente la influencia de los valores a priori obtenidos de las viviendas origen de forma artificial, para comprobar si se puede subsanar o reducir dicho problema de transferencia negativa.

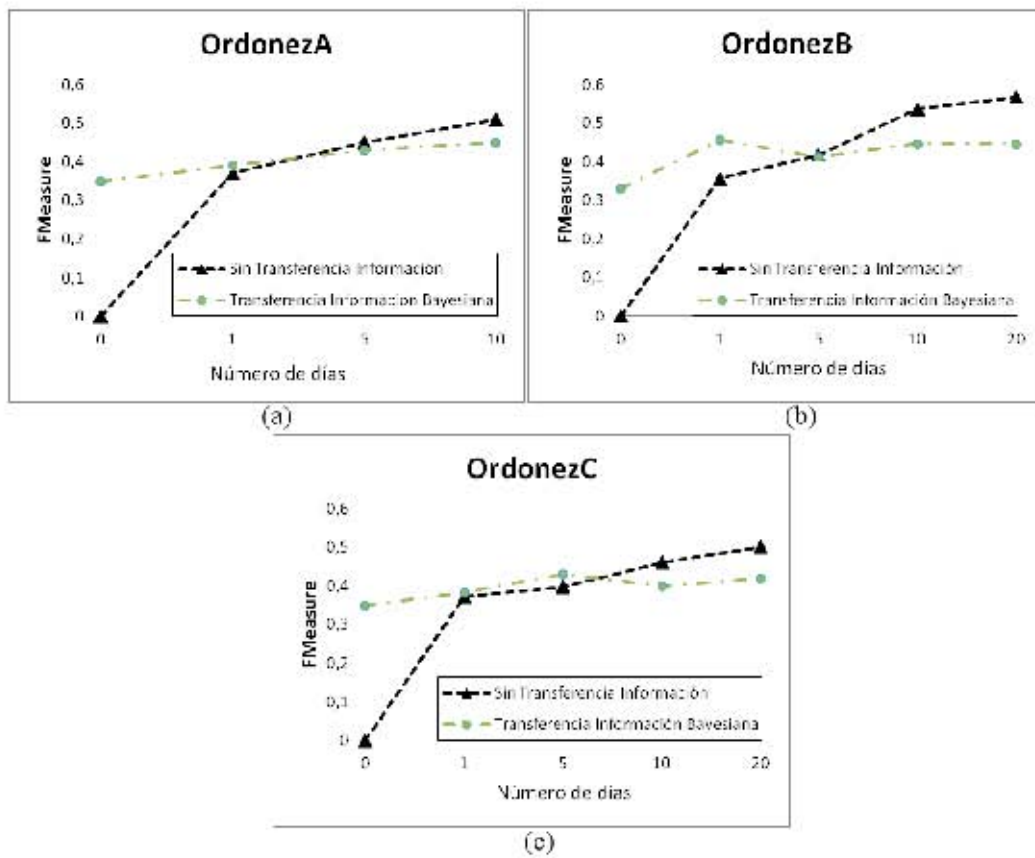


Figura 6.5: Rendimiento de la propuesta para realizar la transferencia de información en la detección de anomalías.

6.3. Conclusiones

En este capítulo hemos presentado un método de transferencia de información y lo hemos evaluado sobre los dos problemas tratados en esta tesis doctoral: el reconocimiento de actividades y la detección de anomalías. En ambos problemas se distingue entre datos objetivo y datos origen, donde los datos objetivo contienen información específica sobre la vivienda en la que se quiere aplicar el método y los datos origen contienen información más genérica. En nuestra propuesta la información es transferida entre viviendas haciendo uso de las distribuciones a priori que modelan los datos, dentro de un enfoque bayesiano.

Cuando se ha de tratar con entornos de monitorización que tienen diversas configuraciones y topologías, normalmente se deben asumir algunas limitaciones para poder integrar datos procedentes de diferentes fuentes. Una solución muy común es considerar que se deben emplear algún tipo de meta características que proporcionen independencia respecto al espacio de datos. En esta propuesta se ha realizado un mapeado sobre el espacio de características original basándose en la función de cada uno de los sensores. El conjunto de meta características empleadas en nuestros experimentos ha sido elegido manualmente y por lo tanto no se puede asegurar que sea una elección óptima. Una alternativa es emplear algún tipo de técnica que pueda definir el meta espacio automáticamente [58, 267].

El transformar el espacio de características supone una pérdida de información para los dos problemas tratados, e implica reducir el rendimiento de los métodos. En el caso del reconocimiento de actividades, esta pérdida es consistente con estudios previos [130], pero sin llegar a ser estadísticamente significativa, lo que implica que la técnica de mapeado permite un rendimiento competitivo y ofrece una solución válida para tratar con las diferentes configuraciones de las redes de sensores. Sin embargo, en el caso de la detección de anomalías, las diferencias son más notables y estadísticamente significativas, dando lugar a un aumento de los falsos negativos, e indicando que el uso de este tipo de mapeo basado en la función de los sensores posiblemente no sea la mejor opción para nuestra propuesta.

El método de transferencia de información para el reconocimiento de actividades presentado supera significativamente el rendimiento de las últimas propuestas probabilísticas en el área. Las dos pruebas de significancia estadística aplicadas muestran cómo nuestro sistema tiene una mayor precisión, en términos de *F-Measure*. Los experimentos demuestran cómo la inferencia bayesiana ofrece mejores resultados para nuestro problema, especialmente cuando se dispone de pocos datos en el conjunto objetivo o cuando se trata con conjuntos de datos que pueden ser ruidosos. La estimación mediante MCMC da lugar a un proceso de inferencia más preciso, y el uso de distribuciones a priori ofrece un mecanismo muy flexible para modelar el espacio de características, permitiendo además un mayor nivel de generalización para poder adaptar diferentes tipos de sensores al

sistema.

Las pruebas realizadas sobre el método de detección de anomalías muestran cómo la transferencia de información resulta útil en este problema cuando se dispone de pocos o ningún dato del conjunto objetivo. No obstante, nuestra propuesta bayesiana es sensible a la transformación del espacio de características, sufriendo una penalización en el rendimiento, e incluso llegando a mostrar transferencia negativa cuando el tamaño del conjunto de datos objetivo aumenta.

Una de las principales limitaciones de los sistemas de modelado del comportamiento es el hecho de que los modelos suelen estar optimizados para una configuración de sensores concreta. Mas allá de los resultados específicos para cada problema, nuestros experimentos confirman que existe la posibilidad de diseñar métodos que permitan abstraer las capacidades de detección de un entorno automáticamente, siendo posible tener algoritmos de modelado que puedan funcionar en meta espacios de características, y de este modo permitan trasladar el aprendizaje desde un entorno instrumentado a otro.

Estas propuestas de transferencia de información permiten avanzar hacia una de las principales metas de la computación ubicua, que es el poder diseñar sistemas que sean capaces de generalizar a través de diferentes dominios.

Capítulo 7

Conclusions

This thesis has approached the problem of modeling human behavior from sensor data for dealing with the societal problems of an aging population. In this final chapter, we present our concluding remarks and highlight the contributions of this thesis to this area of research. We end this chapter with a discussion on directions for future research.

7.1. Conclusions and Contributions

In this thesis we have established several objectives, which represent a number of research questions related with the human behavior modeling problem.

Activities of the Daily Living recognition

First, we focused on addressing the human behavior modeling problem from a supervised learning point of view, using Activity of the Daily Living (ADL) recognition. Given that temporal probabilistic models have been successfully applied in many sequential data modeling problems and have been shown to perform well in this domain, we focused on improving these models. In Chapter 4 we presented two approaches based on hidden Markov models (HMM), namely: hybrid HMM schemes focused to improve the topology of the models and a standard HMM trained by full bayesian learning focused to improve the classical inference process.

Regarding the hybrid HMM schemes, we have proposed the use of two powerful machine learning schemes, ANN (Artificial Neural Network) and SVM (Support Vector Machines), within the framework of HMM in order to tackle the task of ADL recognition. Experimental results of the hybrid HMM models presented demonstrate how different hybrid schemes can be effectively employed for activity recognition in a home setting.

The hybrid schemes presented were compared with a generative model (the

HMM working standalone) and several well known classifiers generated by discriminative algorithms. Results of the experiments showed how purely discriminative models were not able to generalize properly, since they can't model the temporal characteristics of the sequential data. Specifically, we show how the hybrid system obtained by using an SVM to estimate the emission probabilities of an HMM outperforms all other sequential pattern recognition approaches included in the comparative.

By incorporating the time modeling abilities of the HMM to the discriminative skills of the classifier we obtained an efficient scheme that is able to deal with the diverse statistical challenges presented in recognizing human activities and overcome the weakness of HMMs as effective classifiers. Considering the performance of the two hybrid approaches evaluated, our results show how the combination of discriminative and generative models is more accurate than either of the models on their own.

When comparing the presented hybrid approaches with other classifiers in terms of F-measure, hybrid schemes show a significantly better performance, with significance level $p \leq 0,05$, in both Student and Wilcoxon significance tests. Besides, our hybrid models do not require to apply model assumptions and can estimate the emission probabilities with better discriminating properties, increasing the observations space and, without any hypotheses on the statistical distribution of the data, showing how the proposed system is a proper approach to deal with the addressed problem.

However, it must be noticed that hybridizing these schemes implies a more complex system; hence, when integrating into a real home monitoring solution, it should be considered whether performance should take priority over efficiency. Nevertheless, the training phase in a deployed activity recognizer is usually done offline, so we do not consider such growth of complexity a real problem in our domain.

Regarding the standard HMM trained by full bayesian inference, we proved how Markov Chain Monte Carlo (MCMC) techniques can be used to properly estimate the parameters of an HMM in a bayesian configuration. Instead of trying to assign appropriate values to the latent variables of our HMM using expectation-maximization (EM) inference, the bayesian approach we presented sought to identify a distribution over latent variables directly, without ever fixing particular values for the model parameters. To perform inference in our model, we used Gibbs sampling, a stochastic procedure that produces samples from the posterior distribution. The inference method presented allowed to search the space of solutions more completely, overcoming usual drawbacks posed by classical EM algorithms, namely: the use of a recurrent forward-backward algorithm is numerically sensitive, it can get stuck as it approaches a local optima and it can easily overfit the training data. We demonstrated that, for our HMM, taking a bayesian approach to

activity recognition dramatically improves performance over EM estimation. Despite the HMM trained by the MCMC algorithm retained the shortcomings exposed by generative approaches, it offered a much better performance.

When comparing the hybrid schemes with HMMs trained by full bayesian inference, differences in performance were not statistically significant, with significance level $p \leq 0,05$, in both Student and Wilcoxon tests. However, each approach has a different learning mechanism, and this must be taken into account when the system is deployed into a real environment. Hybrid HMM schemes favor the correct classification of more frequent classes over the correct classification of infrequent classes when this would result in misclassification of the frequent class. On the other side, bayesian HMMs tend to misclassify activities with similar values space (e.g. the activities “having breakfast” and “having lunch”).

Anomaly Detection

Another of our objectives was to determine whether unsupervised techniques can be employed to model human behavior under our monitoring system framework. In Chapter 5, we presented an outlier detection method based on bayesian statistics that can be effectively applied for identifying anomalous human behavioral patterns. Such method relies on an approximate estimation of the living patterns of the user, and on prior knowledge that reflects our belief, a priori, of how such patterns should be. We have proposed three new probabilistic features to model the behavior of a person, whose value and efficacy as anomaly detection methods has been illustrated using different datasets collected from instrumented real home environments. These features presented in this thesis provide valuable insights about behavioral patterns of the monitored cohort, and Bayesian statistics has resulted in a very consistent way to reason under the uncertainty of human behavior. Since we use bayesian statistics, our approach easily integrates any known prior knowledge about the living patterns of the user, and new data arriving from the sensors can be incorporated in near real time into the model.

The anomaly detection system is primarily focused on detecting when something unusual is happening in the behaviour of the user, but it is not aimed to yield outliers that could be interesting from a medical point of view. The method can't provide more detail about if an anomaly represents that the health of the user is whether recovering or deteriorating.

Transfer learning

In Chapter 6 we addressed the objective of applying our behavior modeling methods in multiple environments without the need of having training data from each environment. Both activity recognition and anomaly detection models, when trained for one home, can't automatically be used in another home, due to differences in the layout of the places and the behavior of the inhabitants. In this thesis we

proposed a transfer learning approach, which allows us to learn model parameters despite differences between various homes, and also described the importance of accurate transfer learning when automatically modeling the behavior of elderly people. Such transfer learning process required the use of a meta feature representation to map sensor data from datasets of different houses to a common feature space.

When performing transfer learning in the activity recognition domain we have shown that full Bayesian models significantly outperform the state-of-the-art probabilistic models using EM-based MAP. The bayesian approach we advocate in this thesis, in combination with MCMC, enables analysis of all of the parameters and hyperparameters defined for transfer learning. The use of MCMC makes accurate inference possible, and the use of prior distributions provides a very flexible method for modeling the feature space, resulting in a mechanism that can generalize beyond simple binary sensors. Our model showed significantly better performance in terms of F-measure at a confidence level of 95 %, proving Bayesian inference to be better suited for the task. Experiments on five different home settings showed how the performance of the approach is particularly good for a house for which little or no labeled data are available.

When performing transfer learning in the anomaly detection domain, our approach has shown to be very sensitive to transformations in the feature space, even resulting in negative transfer learning in some cases. Nevertheless, the performance of the approach has been good when little or no data were available, showing how anomaly detection can be performed in a transfer learning framework using a meta feature representation.

These experiments on transfer learning showed that it is possible to devise methods that allow to abstract the sensing capabilities of an environment quite automatically, being possible to have behavior modeling systems operating in a meta space that allows the methods to be translated from one sensing configuration to another. This point is very relevant to pervasive computing as the final goal of such area is to design systems that generalize across domains.

Evaluation with real data

Another of our objectives addressed the issue of evaluating the performance of our methods to ensure an accurate performance in a real world setting. The experiments in this thesis were performed on three new real world datasets obtained by our monitoring systems installed in real homes and on three datasets publicly available. Besides, we presented a flexible and cheap sensor system that allows us to create datasets in multiple homes ubiquitously.

Conclusively, the work presented in this thesis further demonstrates that accurate human behavior modeling can be achieved by a set of simple and cheap

state-change sensors installed in a wireless network, using both supervised and unsupervised approaches.

7.2. Future Research

In this thesis, we used wireless sensor networks to capture the behavior of inhabitants. The sensors used were installed in locations that seemed intuitive to the researchers. Performance of our models could be possibly improved by installing additional sensors. However, because these systems will eventually be used on a large scale, it is important to keep the costs of the system to a minimum. It is therefore important to understand the impact on the performance of both the number of sensors used and the location where the sensors are installed. Such understanding will allow us to create behavior modeling systems using the minimum number of sensors needed. A noticeable shortcoming of the monitoring system proposed is the absence of a robust security layer. To meet the necessary privacy and security requirements, some mechanism that permits nodes to have private message exchanges should be included. In future deployment scenarios, the presented monitoring approach will be required to operate in untrusted environments.

In terms of future work for our activity recognizer, further extensions of the hybrid models are feasible, being possible to employ different classifiers as the discriminative layer of our approach, in order to test the performance of different learning paradigms. Also, due to the fact that the hybrid schemes can estimate the emission probabilities with better discriminating properties, it would be valuable to evaluate our approaches with more complex and larger feature spaces.

It would be also valuable to evaluate our system augmenting the data with suitable context information, for instance, adding some time details or including some processed video information. While maintaining the privacy of the users, it can be possible to augment our groundtruth adding some video data, as proposed by [64], using silhouettes instead of raw images, for instance. It has been shown that silhouettes address the privacy concerns of elderly persons and increase their willingness to accept video monitoring systems in their households.

An interesting development would be to employ ensembles of classifiers within the model, as have been proposed in different hybrid approaches. The multifaceted nature of human activities presents opportunities for multiple machine learning schemes to be combined, with the complementary strengths of different approaches meeting different aspects of the computational challenge. Such combination could be applied using either an ensemble of classifiers to estimate the emission probabilities of the Markov chains or specialized classifiers for particular activities.

Regarding the probabilistic features presented to perform anomaly detection, further research can be focused on evaluating the performance of the method when such features are combined in different and more complex ways. Besides, the anomaly detection approach presented should be validated using more datasets in order to ensure the statistical significance of the method.

In future work for the transfer learning problem, the introduction of non-binary sensors is possible within our framework and could bridge the gap from *what* activities are done to *how* they are done. For example, one could imagine measuring how much water is used for washing dishes and inferring how much was washed, and how easily the task was done.

All the experiments performed in this thesis assumed that a household consists of a single person. For a realistic application in a real world setting, a behavior modeling solution should be able to deal households in which multiple people live. The challenge in performing multi-person behavior modeling is that people can perform activities separately, and therefore sensors from different locations would be probably triggered, and activities can be performed jointly, in which a single set of sensors would be employed.

Another future goal is to deploy this system in a real home setting. The models proposed have to be designed to automatically obtain the data, and the output provided could be used to detect deviations in activity patterns and to warn caregivers or relatives of such situation. Such deviations can reflect changes in health status, thus offering the opportunity to apply standard of care diagnostics and to intervene in a timely manner.

A possible scenario is a two level notification system, where different types of alarms are raised depending on the magnitude of the anomaly detected. A small change in the behavior of the user could be notified directly to the family or caregivers, but in case a more severe anomaly is identified, the system could contact directly the social services.

This integration should be done in a convenient and unobtrusive way in order to be acceptable for the user.

7.3. Publications

This section lists the papers where the results of this doctoral thesis have been published:

- JCR Journals:

- Fco. Javier Ordóñez, José Antonio Iglesias, Paula de Toledo, Agapito Ledezma, Araceli Sanchis. “Online activity recognition using evolving classifiers”, *Expert Systems With Applications*, 40 (2013), pp. 1248-1255
 - Fco. Javier Ordóñez, Paula de Toledo, Araceli Sanchis. “Activity Recognition Using Hybrid Generative/Discriminative Models on Home Environments Using Binary Sensors”. *Sensors*, 13, no. 5: 5460-5477.
- International Conferences:
- Andrés Duque, Fco. Javier Ordóñez, Paula de Toledo, and Araceli Sanchis. “Offline and Online Activity Recognition on Mobile Devices using Accelerometer Data”. *IWAAL 2012*: 208-215.
 - Fco. Javier Ordóñez, Andrés Duque, Paula de Toledo, and Araceli Sanchis. “A Hybrid HMM/ANN Model for Activity Recognition in the Home using Binary Sensors”. *IWAAL 2012*: 98-105.
 - José Antonio Iglesias, Fco. Javier Ordóñez, Paula de Toledo, Agapito Ledezma, Araceli Sanchis. “Evolving activity recognition from sensor streams”. *Evolving and Adaptive Intelligent Systems (EAIS)*, 2012 IEEE Conference, pp.96-101, 17-18 May 2012.
- National Conferences:
- Fco. Javier Ordóñez, Andrés Duque, Paula de Toledo, and Araceli Sanchis. “Remote Mobile Monitoring of Wireless Sensor Networks for Ambient Assisted Living”. *CASEIB'12*.

Bibliografía

- [1] Gregory D. Abowd, Anind K. Dey, Peter J. Brown, Nigel Davies, Mark Smith, and Pete Steggles. Towards a better understanding of context and context-awareness. In *Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing*, HUC '99, pages 304–307, London, UK, UK, 1999. Springer-Verlag.
- [2] C Aggarwal. On abnormality detection in spuriously populated data streams. In *Proceedings of 5th SIAM Data Mining*, pages 80–91, 2005.
- [3] Plamen P. Angelov and Xiaowei Zhou. Evolving fuzzy-rule-based classifiers from data streams. *IEEE T. Fuzzy Systems*, 16(6):1462–1475, 2008.
- [4] Fabrizio Angiulli and Clara Pizzuti. Fast outlier detection in high dimensional spaces. In *Principles of Data Mining and Knowledge Discovery*, pages 15–27. Springer, 2002.
- [5] Frank J Anscombe. Rejection of outliers. *Technometrics*, 2(2):123–146, 1960.
- [6] K H Asberg. Assessment of adl in home-care for the elderly. change in adl and use of short-term hospital care. *Scand J Soc Med*, 14(2):105–11, 1986.
- [7] Inc. Assisted Living Services. Care plus. <http://www.assistedlivingct.com/introducing-care-plus.php>, 2013. [Online; visitado 12-Mayo-2013].
- [8] Hagai Attias. A variational bayesian framework for graphical models. *Advances in neural information processing systems*, 12(1-2):209–215, 2000.
- [9] Asier Aztiria, Golnaz Farhadi, and Hamid Aghajan. User behavior shift detection in intelligent environments. In *Ambient Assisted Living and Home Care*, pages 90–97. Springer, 2012.
- [10] V. Babu, L. Prasanth, R. Sharma, G. V. Rao, and A. Bharath. Hmm-based online handwriting recognition system for telugu symbols. In *ICDAR*, pages 63–67. IEEE Computer Society, 2007.

- [11] Ling Bao and Stephen Intille. Activity recognition from user-annotated acceleration data. *Pervasive Computing*, pages 1–17, 2004.
- [12] Vic Barnett and Toby Lewis. Outliers in statistical data. *Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics*, Chichester: Wiley, 1984, 2nd ed., 1, 1984.
- [13] Leonard E Baum, Ted Petrie, George Soules, and Norman Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, 41(1):164–171, 1970.
- [14] Jonathan Baxter. A bayesian/information theoretic model of learning to learn via multiple task sampling. In *Machine Learning*, pages 7–39, 1997.
- [15] Ada-Helen Bayer, Leon Harper, and Mathew Greenwald. *Fixing to stay: A national survey of housing and home modification issues*. American Association of Retired Persons, 2000.
- [16] Yoshua Bengio, Renato De Mori, and Marco Gori. Learning the dynamic nature of speech with back-propagation for sequences. *Pattern recognition letters*, 13(5):375–385, 1992.
- [17] Yoshua Bengio. A connectionist approach to speech recognition. *IJPRAI*, 7(4):647–667, 1993.
- [18] Michael J Berry and Gordon Linoff. *Data mining techniques: for marketing, sales, and customer support*. John Wiley & Sons, Inc., 1997.
- [19] David Berry. *Statistics: A bayesian perspective*. 1996.
- [20] Amiya Bhattacharya and Sajal K Das. Lezi-update: An information-theoretic approach to track mobile users in pcs networks. In *Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking*, pages 1–12. ACM, 1999.
- [21] Christopher M Bishop. *Neural networks for pattern recognition*. Oxford university press, 1995.
- [22] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 1 edition, October 2007.
- [23] Adam Blum. *Neural networks in C++: an object-oriented framework for building connectionist systems*. John Wiley & Sons, Inc., 1992.

- [24] Edwin Bonilla, Kian Ming Chai, and Chris Williams. Multi-task gaussian process prediction. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 153–160. MIT Press, Cambridge, MA, 2008.
- [25] Bruno Bouchard, Sylvain Giroux, and Abdenour Bouzouane. A keyhole plan recognition model for alzheimer’s patients: First results. *Applied Artificial Intelligence*, 21(7):623–658, 2007.
- [26] Herve A. Bourlard and Nelson Morgan. *Connectionist Speech Recognition: A Hybrid Approach*. Kluwer Academic Publishers, Norwell, MA, USA, 1993.
- [27] Hervé Bourlard and Nelson Morgan. Hybrid hmm/ann systems for speech recognition: Overview and new research directions. In *Adaptive Processing of Sequences and Data Structures*, pages 389–417. Springer, 1998.
- [28] Paul C. Boutros and Allan B. Okey. Unsupervised pattern recognition: An introduction to the whys and wherefores of clustering microarray data. *Brief Bioinform*, 6(4):331–343, January 2005.
- [29] Andrew P Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7):1145–1159, 1997.
- [30] R Brause, T Langsdorf, and M Hepp. Neural data mining for credit card fraud detection. In *Tools with Artificial Intelligence, 1999. Proceedings. 11th IEEE International Conference on*, pages 103–106. IEEE, 1999.
- [31] J. Broch, D. Maltz, D. Johnson, Y. C. Hu, and J. Jetcheva. A Performance Comparison of Multihop Wireless Ad Hoc Network Routing Protocols. *Proc. ACM/IEEE Int’l Conf. Mobile Computing and Networking (MobiCom)*, October 1998.
- [32] Rodney A Brooks. The intelligent room project. In *Cognitive Technology, 1997. ‘Humanizing the Information Age’. Proceedings., Second International Conference on*, pages 271–278. IEEE, 1997.
- [33] Suratna Budalakoti, Ashok Srivastava, Ram Akella, and Eugene Turkov. Anomaly detection in large sets of high-dimensional symbol sequences. *NASA Ames Research Center, Tech. Rep. NASA TM-2006-214553*, 2006.
- [34] Michael Buettner, Richa Prasad, Matthai Philipose, and David Wetherall. Recognizing daily activities with rfid-based sensors. In *Proceedings of the 11th international conference on Ubiquitous computing*, Ubicomp ’09, pages 51–60, New York, NY, USA, 2009. ACM.

- [35] Christopher JC Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.
- [36] Rich Caruana. Multitask learning. *Machine Learning*, 28(1):41–75, 1997.
- [37] Rich Caruana. *Multitask learning*. Springer, 1998.
- [38] M Chan, C Hariton, P Ringeard, and E Campo. Smart house automation system for the elderly and the disabled. In *Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century., IEEE International Conference on*, volume 2, pages 1586–1589. IEEE, 1995.
- [39] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3):15, 2009.
- [40] Yin-Wen Chang, Cho-Jui Hsieh, Kai-Wei Chang, Michael Ringgaard, and Chih-Jen Lin. Training and testing low-degree polynomial data mappings via linear svm. *The Journal of Machine Learning Research*, 99:1471–1490, 2010.
- [41] Nitesh V Chawla, Nathalie Japkowicz, and Aleksander Kotcz. Editorial: special issue on learning from imbalanced data sets. *ACM SIGKDD Explorations Newsletter*, 6(1):1–6, 2004.
- [42] Dazhi Chen and Pramod K. Varshney. Qos support in wireless sensor networks: A survey. In *International Conference on Wireless Networks'04*, pages 227–233, 2004.
- [43] Da Chen, Xueguang Shao, Bin Hu, and Qingde Su. Simultaneous wavelength selection and outlier detection in multivariate regression of near-infrared spectra. *Analytical Sciences*, 21(2):161–166, 2005.
- [44] Datong Chen, Jie Yang, and Howard Wactlar. A study of detecting social interaction with sensors in a nursing home environment. *Computer Vision in Human-Computer Interaction*, pages 199–210, 2005.
- [45] Liming Chen, Chris Nugent, Maurice Mulvenna, Dewar Finlay, Xin Hong, and Michael Poland. Using event calculus for behaviour reasoning and assistance in a smart home. In *Proceedings of the 6th international conference on Smart Homes and Health Telematics, ICOST '08*, pages 81–89, Berlin, Heidelberg, 2008. Springer-Verlag.
- [46] Liming Chen, J. Hoey, C.D. Nugent, D.J. Cook, and Zhiwen Yu. Sensor-based activity recognition. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 42(6):790–808, 2012.

- [47] Bill Chiu, Eamonn Keogh, and Stefano Lonardi. Probabilistic discovery of time series motifs. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '03, pages 493–498, New York, NY, USA, 2003. ACM.
- [48] Rodrigo Cilla, Miguel A. Patricio, Jesús García, Antonio Berlanga, and Jose M. Molina. Recognizing human activities from sensors using hidden markov models constructed by feature selection techniques. *Algorithms*, 2(1):282–300, 2009.
- [49] D Cook and M Schmitter-Edgecombe. Activity profiling using pervasive sensing in smart homes. *IEEE Transactions on Information Technology for Biomedicine*, 2008.
- [50] Diane Cook, Kyle D Feuz, and Narayanan C Krishnan. Transfer learning for activity recognition: A survey. *Knowledge and Information Systems*, 2012.
- [51] Diane J. Cook. Health monitoring and assistance to support aging in place. *J. UCS*, 12(1):15–29, 2006.
- [52] Diane J Cook. Learning setting-generalized activity models for smart spaces. *IEEE intelligent systems*, 2010(99):1, 2010.
- [53] Juan M. Corchado, Javier Bajo, Dante I. Tapia, and Ajith Abraham. Using heterogeneous wireless sensor networks in a telemonitoring system for health-care. *Trans. Info. Tech. Biomed.*, 14:234–240, March 2010.
- [54] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.
- [55] Piero Cosi. Hybrid hmm-nn architectures for connected digit recognition. *Neural Networks, IEEE - INNS - ENNS International Joint Conference on*, 5:5085, 2000.
- [56] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21–27, 1967.
- [57] Wenyuan Dai, Gui-Rong Xue, Qiang Yang, and Yong Yu. Transferring naive bayes classifiers for text classification. In *AAAI*, pages 540–545. AAAI Press, 2007.
- [58] Wenyuan Dai, Yuqiang Chen, Gui-Rong Xue, Qiang Yang, and Yong Yu. Translated learning: Transfer learning across different feature spaces. *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*, pages 353–360, 2008.

- [59] Wenyuan Dai, Qiang Yang, Gui-Rong Xue, and Yong Yu. Self-taught clustering. In *Proceedings of the 25th international conference on Machine learning*, ICML '08, pages 200–207, New York, NY, USA, 2008. ACM.
- [60] Gautam Das, King-Ip Lin, Heikki Mannila, Gopal Renganathan, and Padhraic Smyth. Rule discovery from time series. In *KDD*, pages 16–22, 1998.
- [61] Sajal K Das, Diane J Cook, A Battacharya, Edwin O Heierman III, and Tze-Yun Lin. The role of prediction algorithms in the mavhome smart home architecture. *Wireless Communications, IEEE*, 9(6):77–84, 2002.
- [62] Barnan Das, Narayanan C. Krishnan, and Cook J. Moore. Automated activity interventions to assist with activities of daily living. *Agents and Ambient Intelligence*, 2012.
- [63] Raghu Das. Rfid explained. Free IDTech White paper, 2002. Technical report.
- [64] George Demiris, Debra Parker Oliver, Jarod Giger, Marjorie Skubic, and Marilyn Rantz. Older adults' privacy considerations for vision based recognition methods of eldercare applications. *Technology and Health Care*, 17(1):41–48, 2009.
- [65] Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research*, 7:1–30, 2006.
- [66] Thomas G Dietterich. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation*, 10(7):1895–1923, 1998.
- [67] Petar M Djuric and Joon-Hwa Chun. An mcmc sampling approach to estimation of nonstationary hidden markov models. *Signal Processing, IEEE Transactions on*, 50(5):1113–1123, 2002.
- [68] Steve Donoho. Early detection of insider trading in option markets. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 420–429. ACM, 2004.
- [69] Wenliang Du, Lei Fang, and Ning Peng. Lad: Localization anomaly detection for wireless sensor networks. *Journal of parallel and distributed computing*, 66(7):874–886, 2006.
- [70] Thi V. Duong, Hung H. Bui, Dinh Q. Phung, and Svetha Venkatesh. Activity recognition and abnormality detection with the switching hidden semi-markov

- model. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:838–845, 2005.
- [71] T. Duong, D. Phung, H. Bui, and S. Venkatesh. Efficient duration and hierarchical modeling for human activity recognition. *Artificial Intelligence*, 173(7-8):830–856, 2009.
- [72] S. Gong T. Xiang E. Zelniker. Global abnormal behaviour detection using a network of cctv cameras. In *IEEE International Workshop on Visual Surveillance*, 2008.
- [73] R. Emonet, J. Varadarajan, and J. Odobez. Multi-camera open space human activity discovery for anomaly detection. In *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on*, pages 218–223, 30 2011-sept. 2 2011.
- [74] Gwenn Englebiennne. Bayesian methods for the analysis of human behaviour. In *Computer Analysis of Human Behavior*, pages 3–20. 2011.
- [75] S España-Boquera, J Gorbe-Moya, F Zamora-Martínez, and MJ Castro-Bleda. Hybrid hmm/ann models for bimodal online and offline cursive word recognition. In *Recognizing Patterns in Signals, Speech, Images and Videos*, pages 14–21. Springer, 2010.
- [76] Salvador Espana-Boquera, Maria Jose Castro-Bleda, Jorge Gorbe-Moya, and Francisco Zamora-Martinez. Improving offline handwritten text recognition with hybrid hmm/ann models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):767–779, 2011.
- [77] Patrick Th. Eugster, Pascal A. Felber, Rachid Guerraoui, and Anne-Marie Kermarrec. The many faces of publish/subscribe, 2003.
- [78] Michael Evans and Tim Swartz. Methods for approximating integrals in statistics with special emphasis on bayesian integration problems. *Statistical Science*, 10(3):254–272, 1995.
- [79] Ludger Fiege, Gero Mühl, and Felix C. Gärtner. Modular event-based systems. *THE KNOWLEDGE ENGINEERING REVIEW*, 17:359–388, 2006.
- [80] V. Fontaine, C. Ris, and H. Leich. Comparison between two hybrid hmm/mlp approaches in speech recognition. In *Proceedings of the Acoustics, Speech, and Signal Processing, 1996. on Conference Proceedings., 1996 IEEE International Conference - Volume 06, ICASSP '96*, pages 3362–3365, Washington, DC, USA, 1996. IEEE Computer Society.

- [81] Ryohei Fujimaki, Takehisa Yairi, and Kazuo Machida. An approach to spacecraft anomaly detection problem using kernel feature space. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 401–410. ACM, 2005.
- [82] A. Ganapathiraju, J. Hamaker, and J. Picone. Hybrid svm/hmm architectures for speech recognition. In *Speech Transcription Workshop*, pages 504–507, 2000.
- [83] Stuart Geman, Elie Bienenstock, and René Doursat. Neural networks and the bias/variance dilemma. *Neural computation*, 4(1):1–58, 1992.
- [84] Roberto Gemello, Franco Mana, and Dario Albesano. Hybrid hmm/neural network based speech recognition in loquendo asr. URL <http://www.loquendo.com/en/>. [Online], 2010.
- [85] Mohamed F Ghalwash, Dusan Ramljak, and Zoran Obradovic. Early classification of multivariate time series using a hybrid hmm/svm model. In *Bioinformatics and Biomedicine (BIBM), 2012 IEEE International Conference on*, pages 1–6. IEEE, 2012.
- [86] Anthony P Glascock and David M Kutzik. Behavioral telemedicine: A new approach to the continuous nonintrusive monitoring of activities of daily living. *Telemedicine Journal*, 6(1):33–44, 2000.
- [87] Sharon Goldwater and Tom Griffiths. A fully Bayesian approach to unsupervised part-of-speech tagging. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 744–751, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [88] Fabio A González and Dipankar Dasgupta. Anomaly detection using real-valued negative selection. *Genetic Programming and Evolvable Machines*, 4(4):383–403, 2003.
- [89] Subharup Guha, Yi Li, and Donna Neuberg. Bayesian hidden markov modeling of array cgh data. 2006.
- [90] H. Hagiwara, T. Nakano, K. Yoshida, K. Fukuyori, A. Michimori, and E. Koyama. Measurement of human behavior in a daily life based on the understanding of biological rhythm. In *SICE 2002. Proceedings of the 41st SICE Annual Conference*, volume 2, pages 789 – 793 vol.2, aug. 2002.
- [91] Karen Z. Haigh and Holly A. Yanco. Automation as Caregiver: A Survey of Issues and Technologies. In *AAAI Workshop .Automation as Caregiver*, July 2001.

- [92] Rebecca N. Handcock, Dave L. Swain, Greg J. Bishop-Hurley, Kym P. Patison, Tim Wark, Philip Valencia, Peter Corke, and Christopher J. O'Neill. Monitoring animal behaviour and environmental interactions using wireless sensor networks, gps collars and satellite remote sensing. *Sensors*, 9(5):3586–3603, 2009.
- [93] Abhiman Hande, Todd Polk, William Walker, and Dinesh Bhatia. Self-powered wireless sensor networks for remote patient monitoring in hospitals. *Sensors*, 6:1102–1117, 2006.
- [94] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Human mental workload*, 1(3):139–183, 1988.
- [95] Trevor Hastie and Robert Tibshirani. Classification by pairwise coupling. *The annals of statistics*, 26(2):451–471, 1998.
- [96] T. Hastie, R. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning*. Springer, July 2003.
- [97] Simon Haykin and Neural Network. A comprehensive foundation. *Neural Networks*, 2, 2004.
- [98] Simon S Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall Englewood Cliffs, NJ, 2007.
- [99] Healthsense. eNeighbor. <http://healthsense.com/index.php/products/remote-monitoring/eneighbor>, 2013. [Online; visitado 12-Mayo-2013].
- [100] Robert Hecht-Nielsen. Theory of the backpropagation neural network. In *Neural Networks, 1989. IJCNN., International Joint Conference on*, pages 593–605. IEEE, 1989.
- [101] Markus Hegland, William Clarke, and Margaret Kahn. Mining the macho dataset. *Computer Physics Communications*, 142(13):22 – 28, 2001.
- [102] Robert K Heinssen Jr, Carol R Glass, and Luanne A Knight. Assessing computer anxiety: Development and validation of the computer anxiety rating scale. *Computers in human behavior*, 3(1):49–59, 1987.
- [103] Sumi Helal, Bryon Winkler, Choonhwa Lee, Youssef Kaddoura, Lisa Ran, Carlos Giraldo, Sree Kuchibhotla, and William Mann. Enabling location-aware pervasive computing applications for the elderly. In *Pervasive Computing and Communications, 2003.(PerCom 2003). Proceedings of the First IEEE International Conference on*, pages 531–536. IEEE, 2003.

- [104] Sumi Helal, William Mann, Hicham El-Zabadani, Jeffrey King, Youssef Kaddoura, and Erwin Jansen. The gator tech smart house: A programmable pervasive space. *Computer*, 38:50–60, 2005.
- [105] Katherine Heller, Krysta Svore, Angelos D Keromytis, and Salvatore Stolfo. One class support vector machines for detecting anomalous windows registry accesses. In *Workshop on Data Mining for Computer Security (DMSEC), Melbourne, FL, November 19, 2003*, pages 2–9, 2003.
- [106] Jean Hennebert, Christophe Ris, Herve Bourlard, Steve Renals, and Nelson Morgan. Estimation of global posteriors and forward-backward training of hybrid hmm/ann systems. 1997.
- [107] David J. Hill, Barbara S. Minsker, and Eyal Amir. Real-time Bayesian Anomaly Detection for Environmental Sensor Data. In *Proceedings of the 32nd conference of IAHR*, 2007.
- [108] Mike Hochberg, Steve Renals, Anthony J. Robinson, and Dan J. Kershaw. Large vocabulary continuous speech recognition using a hybrid connectionist-hmm system. In *ICSLP. ISCA*, 1994.
- [109] Frank Höppner. Discovery of temporal patterns. learning rules about the qualitative behaviour of time series. In *Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery, PKDD '01*, pages 192–203, London, UK, UK, 2001. Springer-Verlag.
- [110] Qiang Huo, Chorkin Chan, and Chin Hui Lee. Bayesian adaptive learning of the parameters of hidden markov model for speech recognition. *Speech and Audio Processing, IEEE Transactions on*, 3(5):334–345, 1995.
- [111] Tâm Huynh and Bernt Schiele. Unsupervised discovery of structure in activity data using multiple eigenspaces. *Location-and Context-Awareness*, pages 151–167, 2006.
- [112] Tsuyoshi Idé, Spiros Papadimitriou, and Michail Vlachos. Computing correlation anomaly scores using stochastic nearest neighbors. In *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*, pages 523–528. IEEE, 2007.
- [113] Jose Antonio Iglesias, Plamen Angelov, Agapito Ledezma, and Araceli Sanchis. Human activity recognition based on evolving fuzzy systems. *International Journal of Neural Systems*, 20:355–364, 2010.

- [114] Stephen S. Intille, Emmanuel Munguia Tapia, John Rondoni, Jennifer Beaudin, Chuck Kukla, Sitij Agarwal, Ling Bao, and Kent Larson. Tools for studying behavior and technology in natural settings. In *In Proceedings of UBIComp 2003*, pages 157–174. Springer, 2003.
- [115] S. S. Intille, K. Larson, E. Munguia Tapia, J. Beaudin, P. Kaushik, J. Nawyn, and R. Rockinson. Using a live-in laboratory for ubiquitous computing research. In *Proc. of PERVASIVE06*, pages 349–365, 2006.
- [116] Stephen S Intille. Designing a home of the future. *Pervasive Computing, IEEE*, 1(2):76–82, 2002.
- [117] Yoshinori Isoda, Shoji Kurakake, and Hirotaka Nakano. Ubiquitous sensors based human behavior modeling and recognition using a spatio-temporal representation of user states. In *Advanced Information Networking and Applications, 2004. AINA 2004. 18th International Conference on*, volume 1, pages 512–517. IEEE, 2004.
- [118] Tommi S Jaakkola, David Haussler, et al. Exploiting generative models in discriminative classifiers. *Advances in neural information processing systems*, pages 487–493, 1999.
- [119] Roozbeh Jafari, Andre Encarnacao, Azad Zahoory, Foad Dabiri, Hyduke Noshadi, and Majid Sarrafzadeh. Wireless sensor networks for health monitoring. In *MobiQuitous*, pages 479–781, 2005.
- [120] D Janakiram, V Adi Mallikarjuna Reddy, and AVU Phani Kumar. Outlier detection in wireless sensor networks using bayesian belief networks. In *Communication System Software and Middleware, 2006. Comsware 2006. First International Conference on*, pages 1–6. IEEE, 2006.
- [121] Nathalie Japkowicz and Shaju Stephen. The class imbalance problem: A systematic study. *Intell. Data Anal.*, 6(5):429–449, 2002.
- [122] Mark Johnson. Why doesnt em find good hmm pos-taggers. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 296–305, 2007.
- [123] A Jordan. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. *Advances in neural information processing systems*, 14:841, 2002.

- [124] Anagha Joshi, Yves Van de Peer, and Tom Michoel. Analysis of a gibbs sampler method for model-based clustering of gene expression data. *Bioinformatics*, 24(2):176–183, January 2008.
- [125] T. L. M. Kasteren. Datasets for activity recognition. <http://sites.google.com/site/tim0306/>. [Online; visitado 24-November-2012].
- [126] T. L. M. Kasteren, G. Englebienne, and B. J. A. Kröse. Recognizing activities in multiple contexts using transfer learning. In *Proceedings of the AAAI Fall Symposium on AI in Eldercare: New Solutions to Old Problems*. AAAI Press, 2008. ISBN=978-1-57735-394-2.
- [127] T. L. M. Kasteren, A. Noulas, G. Englebienne, and B. J. A. Kröse. Accurate activity recognition in a home setting. In *Proceedings of the 10th international conference on Ubiquitous computing*, UbiComp '08, pages 1–9, New York, NY, USA, 2008. ACM.
- [128] T. L. M. Kasteren, G. Englebienne, and B. J. A. Kröse. An activity monitoring system for elderly care using generative and discriminative models. *Personal Ubiquitous Comput.*, 14:489–498, September 2010.
- [129] T. L. M. Kasteren, G. Englebienne, and B. J. A. Kröse. Activity recognition using semi-markov models on real world smart home datasets. *Journal of ambient intelligence and smart environments*, 2(3):311–325, 2010.
- [130] T. L. M. Kasteren, G. Englebienne, and B. J. A. Kröse. Transferring knowledge of activity recognition across sensor networks. In *Pervasive*, pages 283–300, 2010.
- [131] T. L. M. Kasteren, G. Englebienne, and B. J. A. Kröse. Human Activity Recognition from Wireless Sensor Network Data: Benchmark and Software. In Liming Chen, Chris D. Nugent, Jit Biswas, and Jesse Hoey, editors, *Activity Recognition in Pervasive Intelligent Environments*, volume 4 of *Atlantis Ambient and Pervasive Intelligence*, chapter 8, pages 165–186. Atlantis Press, Paris, France, 2011.
- [132] T. L. M. Kasteren. *Activity recognition for health monitoring elderly using temporal probabilistic models*. PhD thesis, University of Amsterdam, 2011.
- [133] S. Katz. Assessing self-maintenance: Activities of daily living, mobility, and instrumental activities of daily living. *Journal of the American Geriatrics Society*, 31(12):721, 1983.

- [134] Henry A Kautz. *A formal theory of plan recognition*. PhD thesis, Bell Laboratories, 1987.
- [135] Eamonn Keogh and M. Pazzani. An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback. In R. Agrawal, P. Stolorz, and G. Piatetsky-Shapiro, editors, *Proceedings of 4th KDD*, pages 239–241, New York City, NY, 1998. ACM Press.
- [136] Cory Kidd, Robert Orr, Gregory Abowd, Christopher Atkeson, Irfan Essa, Blair MacIntyre, Elizabeth Mynatt, Thad Starner, and Wendy Newstetter. The Aware Home: A Living Laboratory for Ubiquitous Computing Research. In Norbert Streitz, Jane Siegel, Volker Hartkopf, and Shin’ Konomi, editors, *Cooperative Buildings. Integrating Information, Organizations and Architecture*, volume 1670 of *Lecture Notes in Computer Science*, chapter 17, pages 191–198. Springer Berlin / Heidelberg, Berlin, Heidelberg, 1999.
- [137] Josef Kittler, Mohamad Hatef, Robert P. W. Duin, and Jiri Matas. On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(3):226–239, 1998.
- [138] Ron Kohavi. The power of decision tables. In *8th European Conference on Machine Learning*, pages 174–189. Springer, 1995.
- [139] H-P. Kriegel, P. Kröger, and A. Zimek. Outlier detection techniques. In *16th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, 2010.
- [140] Jennifer R. Kwapisz, Gary M. Weiss, and Samuel A. Moore. Activity recognition using cell phone accelerometers. *SIGKDD Explor. Newsl.*, 12(2):74–82, March 2011.
- [141] James Landay and Anthony LaMarca. Intels seattle lab making an impact. Intel, 2005. Technical report.
- [142] M. P. Lawton and E. M. Brody. Assessment of older people: self-maintaining and instrumental activities of daily living. *The Gerontologist*, 9(3):179–186, 1969.
- [143] Aleksandar Lazarevic, Levent Ertoz, Vipin Kumar, Aysel Ozgur, and Jai-deep Srivastava. A comparative study of anomaly detection schemes in network intrusion detection. In *Proceedings of the third SIAM international conference on data mining*, volume 3, pages 25–36. Siam, 2003.

- [144] Lilia Lazli, Abdennasser Chebira, Mohamed Tayeb Laskri, and Kurosh Madani. Hybrid hmm/ann system using fuzzy clustering for speech and medical pattern recognition. In *Digital Information and Communication Technology and Its Applications*, pages 557–570. Springer, 2011.
- [145] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y. Ng. Efficient sparse coding algorithms. In *In NIPS*, pages 801–808. NIPS, 2007.
- [146] S. I Lee, V. Chatalbashev, D. Vickrey, and D. Koller. Learning a meta-level prior for feature relevance from multiple related tasks. In *Proceedings of the 24th international conference on Machine learning*, page 489–496, 2007.
- [147] Jonathan Lester, Tanzeem Choudhury, Nicky Kern, Gaetano Borriello, and Blake Hannaford. A hybrid discriminative/generative approach for modeling human activities. In *In Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 766–772, 2005.
- [148] Lin Liao, Dieter Fox, and Henry Kautz. Location-based activity recognition using relational markov networks. In *Proceedings of the 19th international joint conference on Artificial intelligence, IJCAI'05*, pages 773–778, San Francisco, CA, USA, 2005. Morgan Kaufmann Publishers Inc.
- [149] Jessica Lin, Eamonn Keogh, Ada Fu, and Helga Van Herle. Approximations to magic: Finding unusual medical time series. In *Computer-Based Medical Systems, 2005. Proceedings. 18th IEEE Symposium on*, pages 329–334. IEEE, 2005.
- [150] Jun S Liu. *Monte Carlo strategies in scientific computing*. Springer Verlag, 2008.
- [151] Beth Logan, Jennifer Healey, Matthai Philipose, Emmanuel Munguia Tapia, and Stephen Intille. A long-term evaluation of sensing modalities for activity recognition. In *Proceedings of the 9th international conference on Ubiquitous computing, UbiComp '07*, pages 483–500, Berlin, Heidelberg, 2007. Springer-Verlag.
- [152] Paul Lukowicz, Holger Junker, Mathias Stäger, T. von Büren, and Gerhard Tröster. Wearnnet: A distributed multi-sensor system for context aware wearables. In Gaetano Borriello and Lars Erik Holmquist, editors, *Ubicomp*, volume 2498 of *Lecture Notes in Computer Science*, pages 361–370. Springer, 2002.
- [153] M. S. Magnusson. Discovering hidden time patterns in behavior: T-patterns and their detection. *Behav Res Methods Instrum Comput*, 32(1):93–110, February 2000.

- [154] Dominic Maguire and Richard Frisby. Comparison of feature classification algorithm for activity recognition based on accelerometer and heart rate data. In *9th. IT & T Conference*, page 11, 2009.
- [155] Dimitrios Makris and Tim Ellis. Path detection in video surveillance. *Image and Vision Computing*, 20:895–903, 2002.
- [156] Steve Mann. Definition of wearable computing. In *Proceedings of the 1998 International Conference on Wearable Computing*, 1998.
- [157] Konstantin Markov and Satoshi Nakamura. Using hybrid hmm/bn acoustic models: Design and implementation issues. *IEICE - Trans. Inf. Syst.*, E89-D(3):981–988, March 2006.
- [158] Konstantin Markov, Jianwu Dang, Yosuke Iizuka, and Satoshi Nakamura. Hybrid hmm/bn asr system integrating spectrum and articulatory features. In *Proc. Eurospeech*, pages 965–968, 2003.
- [159] S. Marukatat, T. Artires, P. Gallinari, and B. Dorizzi. Sentence recognition through hybrid neuro-markovian modeling. In *In International Conference on Document Analysis and Recognition (ICDAR)*, pages 731–735, 2001.
- [160] Uwe Maurer, Anthony Rowe, Asim Smailagic, and Daniel Siewiorek. Location and activity recognition using ewatch: A wearable sensor platform. *Ambient Intelligence in Everyday Life*, pages 86–102, 2006.
- [161] Uwe Maurer, Asim Smailagic, Daniel P. Siewiorek, and Michael Deisher. Activity recognition and monitoring using multiple sensors on different body positions. In *Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks*, pages 113–116. IEEE Computer Society, 2006.
- [162] John McCarthy. Circumscription - a form of non-monotonic reasoning. *Artif. Intell.*, 13(1-2):27–39, 1980.
- [163] Ross Messing, Chris Pal, and Henry Kautz. Activity recognition using the velocity histories of tracked keypoints. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 104–111. IEEE, 2009.
- [164] Ar Milenkovic, Chris Otto, and Emil Jovanov. Wireless sensor networks for personal health monitoring: Issues and an implementation. *Computer Communications*, 29:2521–2533, 2006.

- [165] David Minnen, Thad Starner, Irfan Essa, and Charles Isbell. Discovering characteristic actions from on-body sensor data. In *In Proc. of IEEE International Symposium on Wearable Computing*, pages 11–18, 2006.
- [166] Muhamad Azman Miskam, Azwan B. Nasirudin, and Inzarulfaisham A. Rahim. Preliminary design on the development of wireless sensor network for paddy rice cropping monitoring application in malaysia. *European Journal of Scientific Research*, 37:649–657, 2009.
- [167] Frank G Miskelly. Assistive technology in elderly care. *Age and ageing*, 30(6):455–458, 2001.
- [168] Thomas B. Moeslund, Adrian Hilton, and Volker Krüger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2-3):90–126, November 2006.
- [169] Anuj Mohamed and KN Nair. Hmm/ann hybrid model for continuous malayalam speech recognition. *Procedia Engineering*, 30:616–622, 2012.
- [170] Michael C Mozer, Robert H Dodier, Marc Anderson, Lucky Vidmar, RF Cruickshank, and Debra Miller. The neural network house: an overview. *Current trends in connectionism*, pages 371–380, 1995.
- [171] MQ Telemetry Transport. <http://mqtt.org/>. [Online; visitado 25-July-2011].
- [172] Hammadi Nait-Charif and Stephen J. McKenna. Activity summarisation and fall detection in a supportive home environment. In *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 4 - Volume 04*, ICPR '04, pages 323–326, Washington, DC, USA, 2004. IEEE Computer Society.
- [173] Futoshi Naya, Ren Ohmura, Fusako Takayanagi, Haruo Noma, and Kiyoshi Kogure. Workers Routine Activity Recognition using Body Movements and Location Information. In *Wearable Computers, 2006 10th IEEE International Symposium on*, pages 105–108, 2006.
- [174] Darfiana Nur, David Allingham, Judith Rousseau, Kerrie L. Mengersen, and Ross McVinish. Bayesian hidden markov model for dna sequence segmentation: A prior sensitivity analysis. *Computational Statistics Data Analysis*, 53(5):1873 – 1882, 2009.
- [175] Koninklijke Philips Electronics N.V. Philips lifeline - medical alert service. <http://www.lifelinesys.com>, 2013. [Online; visitado 12-Mayo-2013].

- [176] Tim Oates, Matthew D. Schmill, and Paul R. Cohen. A method for clustering the experiences of a mobile robot that accords with human judgments. In *in Proceedings of IJCAI*, pages 846–851, 2000.
- [177] T. Oates. PERUSE: An unsupervised algorithm for finding recurring patterns in time series. In *Data Mining, 2002. ICDM 2002. Proceedings. 2002 IEEE International Conference on*, pages 330–337, 2002.
- [178] S. Ohta, H. Nakamoto, Y. Shinagawa, and T. Tanikawa. A health monitoring system for elderly people living alone. *Journal of Telemedicine and Telecare*, 8(3):151–156, 2002. Cited By (since 1996): 31Export Date: 14 September 2011Source: Scopus.
- [179] Daniel Olguín Olguín and Alex Pentland. Human activity recognition: Accuracy across common locations for wearable sensors. In *2006 10th IEEE International Symposium on Wearable Computers*, pages 11–13, 2006.
- [180] W. H. Organization. Global age-friendly cities, 2007.
- [181] Adithyan Palaniappan, R Bhargavi, and V Vaidehi. Abnormal human activity recognition using svm based approach. In *Recent Trends In Information Technology (ICRTIT), 2012 International Conference on*, pages 97–102. IEEE, 2012.
- [182] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, October 2010.
- [183] Rita Paradiso and Danilo De Rossi. Advances in textile technologies for unobtrusive monitoring of vital parameters and movements. *Conf Proc IEEE Eng Med Biol Soc*, 1:392–5, 2006.
- [184] Donald Patterson, Lin Liao, Dieter Fox, and Henry Kautz. Inferring high-level behavior from low-level sensors. In *UbiComp 2003: Ubiquitous Computing*, pages 73–89. Springer, 2003.
- [185] D. J Patterson, D. Fox, H. Kautz, and M. Philipose. Fine-grained activity recognition by aggregating abstract object usage. In *Wearable Computers, 2005. Proceedings. Ninth IEEE International Symposium on*, page 44–51, 2005.
- [186] Matthai Philipose, Kenneth P. Fishkin, Mike Perkowitz, Donald J. Patterson, Dieter Fox, Henry Kautz, and Dirk Hahnel. Inferring activities from interactions with objects. *IEEE Pervasive Computing*, 3(4):50–57, October 2004.

- [187] Clifton Phua, Daminda Alahakoon, and Vincent Lee. Minority report in fraud detection: classification of skewed data. *ACM SIGKDD Explorations Newsletter*, 6(1):50–59, 2004.
- [188] Antonio Pievatolo, Fabrizio Ruggeri, and Refik Soyer. A bayesian hidden markov model for imperfect debugging. *Reliability Engineering & System Safety*, 2012.
- [189] Susanna Pirttikangas, Kaori Fujinami, and Tatsuo Nakajima. Feature selection and activity recognition from wearable sensors. In *UCS*, pages 516–527, 2006.
- [190] John Platt et al. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3):61–74, 1999.
- [191] Martha E. Pollack. Intelligent technology for an aging population the use of ai to assist elders with cognitive impairment. *AI Magazine*, 26(2):9–24, 2005.
- [192] LLC pomdevices. Sonamba. /<http://sonamba.com/>, 2013. [Online; visitado 12-Mayo-2013].
- [193] R. W. Poppe. A survey on vision-based human action recognition. *Image and Vision Computing*, 28(6):976–990, June 2010.
- [194] Pere Pujol, Susagna Pol, Climent Nadeu, Astrid Hagen, and Hervé Bourlard. Comparison and combination of features in a hybrid hmm/mlp and a hmm/gmm speech recognition system. *Speech and Audio Processing, IEEE Transactions on*, 13(1):14–22, 2005.
- [195] J. Pynoos. Neglected areas in gerontology: Housing adaptation, 2002.
- [196] J. Ross Quinlan. *C4.5: Programs for Machine Learning (Morgan Kaufmann Series in Machine Learning)*. Morgan Kaufmann, 1 edition, October 1992.
- [197] Lawrence R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [198] Kambiz Rahbar and Ali Broumandnia. Independent-speaker isolated word speech recognition based on mean-shift framing using hybrid hmm/svm classifier. In *Electrical Engineering (ICEE), 2010 18th Iranian Conference on*, pages 156–161. IEEE, 2010.

- [199] Rajat Raina, Andrew Y. Ng, and Daphne Koller. Constructing informative priors using transfer learning. In *Proceedings of the 23rd international conference on Machine learning*, ICML '06, pages 713–720, New York, NY, USA, 2006. ACM.
- [200] Rajat Raina, Alexis Battle, Honglak Lee, Benjamin Packer, and Andrew Y. Ng. Self-taught learning: transfer learning from unlabeled data. In *Proceedings of the 24th international conference on Machine learning*, ICML '07, pages 759–766, New York, NY, USA, 2007. ACM.
- [201] Nishkam Ravi, Nikhil Dandekar, Preetham Mysore, and Michael L Littman. Activity recognition from accelerometer data. In *Proceedings of the national conference on artificial intelligence*, volume 20, page 1541. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2005.
- [202] Stephan Reiter, Björn Schuller, and Gerhard Rigoll. Segmentation and recognition of meeting events using a two-layered hmm and a combined mlp-hmm approach. In *Multimedia and Expo, 2006 IEEE International Conference on*, pages 953–956. IEEE, 2006.
- [203] S. Renals, N. Morgan, H. Bourlard, M. Cohen, and H. Franco. Connectionist probability estimators in hmm speech recognition. *Speech and Audio Processing, IEEE Transactions on*, 2(1):161–174, jan 1994.
- [204] RFM. [online; visitado 25-july-2012]. <http://www.rfm.com>, 2012.
- [205] Michael D Richard and Richard P Lippmann. Neural network classifiers estimate bayesian a posteriori probabilities. *Neural computation*, 3(4):461–483, 1991.
- [206] G Rigoil and Christoph Neukirchen. A new approach to hybrid hmm/ann speech recognition using mutual information neural networks. *Advances in Neural Information Processing Systems*, pages 772–778, 1997.
- [207] Irina Rish. An empirical study of the naive bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, pages 41–46, 2001.
- [208] Christian P. Robert, Gilles Celeux, and Jean Diebolt. Bayesian estimation of hidden markov chains: a stochastic implementation. *Statistics Probability Letters*, 16(1):77–83, 1993.
- [209] Christian P Robert, Tobias Ryden, and David M Titterington. Bayesian inference in hidden markov models through the reversible jump markov chain

- monte carlo method. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1):57–75, 2000.
- [210] Y Dan Rubinstein, Trevor Hastie, et al. Discriminative vs informative learning. In *Proc. Third Int. Conf. on Knowledge Discovery and Data Mining*, pages 49–53, 1997.
- [211] Tobias Ryden. EM versus Markov chain Monte Carlo for Estimation of Hidden Markov Models: A Computational Perspective. *Bayesian Analysis*, 2008.
- [212] Joseph Rynkiewicz. Hybrid hmm/mlp models for time series prediction. In *ESANN99*, pages 455–462, 1999.
- [213] Daniel Salber, Anind K. Dey, Rob J. Orr, and Gregory D. Abowd. Designing for ubiquitous computing: A case study in context sensing. Technical report, Usability Center, jul. 1999.
- [214] Steven L Scott. Bayesian methods for hidden markov models : Recursive computing in the 21st century. *Journal of the American Statistical Association*, 97(457):337–351, 2002.
- [215] Andrew W. Senior and Anthony J. Robinson. An off-line cursive handwriting recognition system. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(3):309–321, 1998.
- [216] Jae Hyuk Shin, Boreom Lee, and Kwang Suk Park. Detection of abnormal living patterns for elderly living alone using support vector data description. *Information Technology in Biomedicine, IEEE Transactions on*, 15(3):438 – 448, may 2011.
- [217] Katayoun Sohrabi, Jay Gao, Vishal Ailawadhi, and Gregory J Pottie. Protocols for self-organization of a wireless sensor network. *IEEE Personal Communications*, 7:16–27, 2000.
- [218] Jan Stadermann and Gerhard Rigoll. A hybrid svm/hmm acoustic modeling approach to automatic speech recognition. In *INTERSPEECH*, 2004.
- [219] H Brooke Stauffer. Smart enabling system for home automation. *Consumer Electronics, IEEE Transactions on*, 37(2):xxix–xxxv, 1991.
- [220] Maja Stikic, Tâm Huynh, Kristof Van Laerhoven, and Bernt Schiele. Adl recognition based on the combination of rfid and accelerometer sensing. In *Proceedings of the 2nd International Conference on Pervasive Computing Techno-*

- logies for Healthcare (Pervasive Health 2008)*, page 258–263, Tampere, Finland, January 2008. IEEE Xplore, IEEE Xplore.
- [221] Yang Sun, Louis Ten Bosch, and Lou Boves. Hybrid hmm/blstm-rnn for robust speech recognition. In *Text, Speech and Dialogue*, pages 400–407. Springer, 2010.
- [222] N.K. Suryadevara, A. Gaddam, S.C. Mukhopadhyay, and R.K. Rayudu. Wellness determination of inhabitant based on daily activity behaviour in real-time monitoring using sensor networks. In *Sensing Technology (ICST), 2011 Fifth International Conference on*, pages 474–481, 28 2011-dec. 1 2011.
- [223] Ryoji Suzuki, Mitsuhiro Ogawa, Sakuko Otake, Takeshi Izutsu, Yoshiko Tobimatsu, Tsutomu Iwaya, and Shin-Ichi Izumi. Rhythm of daily living and detection of atypical days for elderly people living alone as determined with a monitoring system. *J Telemed Telecare*, 12(4):208–214, June 2006.
- [224] GrandCare Systems. <http://www.grandcare.com/>, 2013. [Online; visitado 12-Mayo-2013].
- [225] S. Szewczyk, K. Dwan, B. Minor, B. Swedlove, and D. Cook. Annotating smart environment sensor data for activity learning. *Technol. Health Care*, 17(3):161–169, August 2009.
- [226] T. Tam, A. Dolan, J. Boger, and A. Mihailidis. An intelligent emergency response system: Preliminary development and testing of a functional health monitoring system. *Gerontechnology*, 4(4), 2006.
- [227] Gaurav Tandon and Philip K Chan. Weighting versus pruning in rule validation for detecting network and host anomalies. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 697–706. ACM, 2007.
- [228] Puay Tang and Tim Venables. 'smart'homes and telecare for independent living. *Journal of Telemedicine and Telecare*, 6(1):8–14, 2000.
- [229] Martin A Tanner. *Tools for statistical inference: observed data and data augmentation methods*. Springer-Verlag New York, 1991.
- [230] Emmanuel Munguia Tapia, Stephen S. Intille, and Kent Larson. Activity recognition in the home using simple and ubiquitous sensors. In *Pervasive*, pages 158–175, 2004.
- [231] Romain Tavenard, Albert Ali Salah, and Eric Pauwels. Searching for temporal patterns in ami sensor data. In *AI Methods for AmI*, 2007.

- [232] Odin Taylor and D Addison. Novelty detection using neural network technology. In *COMADEM 2000: 13 th International Congress on Condition Monitoring and Diagnostic Engineering Management*, pages 731–743, 2000.
- [233] Sebastian Thrun. Is learning the n-th thing any easier than learning the first? In *Advances in Neural Information Processing Systems*, pages 640–646. The MIT Press, 1996.
- [234] Luke Tierney. Markov chains for exploring posterior distributions. *Annals of Statistics*, 22:1701–1762, 1994.
- [235] Antonio Torralba, Kevin P Murphy, and William T Freeman. Sharing features: efficient boosting procedures for multiclass object detection. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–762. IEEE, 2004.
- [236] Edmondo Trentin and Marco Gori. A survey of hybrid ann/hmm models for automatic speech recognition. *Neurocomputing*, 37(1):91–126, 2001.
- [237] Edmondo Trentin and Marco Gori. Robust combination of neural networks and hidden markov models for speech recognition. *Neural Networks, IEEE Transactions on*, 14(6):1519–1531, 2003.
- [238] William MK Trochim. Likert scaling. *Research methods knowledge base*, 2, 2006.
- [239] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea. Machine Recognition of Human Activities: A Survey. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(11):1473–1488, September 2008.
- [240] Tijs van Dam and Koen Langendoen. An adaptive energy-efficient mac protocol for wireless sensor networks. In *Proceedings of the 1st international conference on Embedded networked sensor systems*, SenSys '03, pages 171–180, New York, NY, USA, 2003. ACM.
- [241] Tran Van Phuong, Seong Jin Cho, Young-Koo Lee, Sungyoung Lee, et al. An anomaly detection algorithm for detecting attacks in wireless sensor networks. In *Intelligence and Security Informatics*, pages 735–736. Springer, 2006.
- [242] Vladimir Vapnik. *The nature of statistical learning theory*. springer, 1999.
- [243] Dan Ventura and Sean Warnick. A theoretical foundation for inductive transfer. Brigham Young University, College of Physical and Mathematical Sciences, 2007. (Abstract Only).

- [244] Adam Vinuela and G Grudic. Unsupervised outlier detection and semi-supervised learning. Technical report, Tech. rep. CU-CS-976-04, University of Colorado at Boulder, 2004.
- [245] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511. IEEE, 2001.
- [246] Gilles Virone, Norbert Noury, and Jacques Demongeot. A system for automatic measurement of circadian activity deviations in telemedicine. *Biomedical Engineering, IEEE Transactions on*, 49(12):1463–1469, 2002.
- [247] Gilles Virone, Majd Alwan, Siddharth Dalal, Steven W. Kell, Beverly Turner, John A. Stankovic, and Robin A. Felder. Behavioral patterns of older adults in assisted living. *IEEE Transactions on Information Technology in Biomedicine*, 12(3):387–398, 2008.
- [248] Brian Walsh. Markov chain monte carlo and gibbs sampling. 2004.
- [249] Jack M. Wang, David J. Fleet, and Aaron Hertzmann. Gaussian process dynamical models for human motion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(2):283–298, February 2008.
- [250] Zheng Wang, Yangqiu Song, and Changshui Zhang. Transferred dimensionality reduction. In *Proceedings of the European conference on Machine Learning and Knowledge Discovery in Databases - Part II, ECML PKDD '08*, pages 550–565, Berlin, Heidelberg, 2008. Springer-Verlag.
- [251] G. Williams, K. Doughty, and D. A. Bradley. A systems approach to achieving carernet-an integrated and intelligent telecare system. *Trans. Info. Tech. Biomed.*, 2(1):1–9, March 1998.
- [252] Daniel H Wilson and Chris Atkeson. Simultaneous tracking and activity recognition (star) using many anonymous, binary sensors. In *Pervasive computing*, pages 62–79. Springer, 2005.
- [253] Daniel H Wilson. *Assistive Intelligent Environments for Automatic In-Home Health Monitoring*. PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, September 2005.
- [254] K.-J. Won, A. Prugel-Bennett, and A. Krogh. Training hmm structure with genetic algorithm for biological sequence analysis. *Bioinformatics*, 20:3613–3619, December 2004.

- [255] Weng-Keen Wong, Andrew Moore, Gregory Cooper, and Michael Wagner. Bayesian network anomaly pattern detection for disease outbreaks. In *ICML*, pages 808–815, 2003.
- [256] Christopher Richard Wren and Emmanuel Munguia Tapia. Toward scalable activity recognition for sensor networks. In *LoCA*, volume 3987 of *Lecture Notes in Computer Science*, pages 168–185. Springer, 2006.
- [257] Xinyu Wu, Yongsheng Ou, Huihuan Qian, and Yangsheng Xu. A detection system for human abnormal behavior. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 1204–1208. IEEE, 2005.
- [258] Jianxin Wu, Adebola Osuntogun, Tanzeem Choudhury, Matthai Philipose, and James M. Rehg. A scalable approach to activity recognition based on object use. In *In Proceedings of the International Conference on Computer Vision (ICCV)*, 2007.
- [259] Tao Xiang and Shaogang Gong. Video behavior profiling for anomaly detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(5):893–908, 2008.
- [260] A. Yamaguchi, M. Ogawa, T. Tamura, and T. Togawa. Monitoring behavior in the home using positioning sensors. In *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*, volume 4, pages 1977 –1979 vol.4, oct-1 nov 1998.
- [261] Junji Yamato, Jun Ohya, and Kenichiro Ishii. Recognizing human action in time-sequential images using hidden markov model. In *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on*, pages 379–385. IEEE, 1992.
- [262] C Yau, Omiros Papaspiliopoulos, Gareth O Roberts, and Christopher Holmes. Bayesian non-parametric hidden markov models with applications in genomics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(1):37–57, 2011.
- [263] Jie Yin, Qiang Yang, and Jeffrey Junfeng Pan. Sensor-based abnormal human-activity detection. *Knowledge and Data Engineering, IEEE Transactions on*, 20(8):1082–1090, 2008.
- [264] Piero Zappi, Clemens Lombriser, Thomas Stiefmeier, Elisabetta Farella, Daniel Roggen, Luca Benini, and Gerhard Tröster. Activity recognition from on-body sensors: Accuracy-power trade-off by dynamic sensor selection. In

- European Conference on Wireless Sensor Networks*, pages 17–33. Springer-Verlag, 2008.
- [265] Lidan Zhang and Kwop-Ping Chan. Adaptive bayesian hmm for fully unsupervised chinese part-of-speech induction. *ACM Transactions on Asian Language Information Processing (TALIP)*, 11(3):9, 2012.
- [266] Ji Zhang and Hai Wang. Detecting outlying subspaces for high-dimensional data: the new task, algorithms, and performance. *Knowledge and information systems*, 10(3):333–355, 2006.
- [267] Jian Zhang, Zoubin Ghahramani, and Yiming Yang. Learning multiple related tasks using latent independent component analysis. *Advances in neural information processing systems*, 18:1585, 2006.
- [268] Kejia Zhang, Shengfei Shi, Hong Gao, and Jianzhong Li. Unsupervised outlier detection in sensor networks using aggregation tree. In *Advanced Data Mining and Applications*, pages 158–169. Springer, 2007.
- [269] Irving Kenneth Zola. *Living at Home: The convergence of aging and disability*, pages 25–39. Baywood Publishing, Amityville, NY, 1997. Print retained.

Apéndice A

Introducción a los Modelos Ocultos de Markov

Los modelos ocultos de Markov (HMM del inglés *Hidden Markov Models*) fueron presentados por Baum y Petrie en el año 1966 [13].

Este anexo es una introducción general a los modelos ocultos de Markov. En él se explican las nociones básicas necesarias para la comprensión del modo en que se han usado los HMM en esta tesis doctoral.

A.1. Definición de *Modelos Ocultos de Markov*

Los procesos de *Markov* son modelos matemáticos que describen el comportamiento de un sistema dinámico sometido a un fenómeno de naturaleza aleatoria. Este tipo de procesos son conocidos como procesos estocásticos y tienen gran importancia en el análisis y estudio de sistemas dinámicos.

Una cadena de Markov, también conocida como modelo de Markov observado, es una secuencia de eventos donde la probabilidad de cada uno de ellos depende exclusivamente del evento precedente. Así, una cadena de Markov puede considerarse como un tipo de modelo gráfico probabilístico cuyos estados son observables.

Un HMM es un conjunto finito de estados relacionados mediante una probabilidad de distribución. La transición entre los estados está definida por una serie de probabilidades denominadas *probabilidades de transición*. Pero en este caso, un HMM tiene dos procesos estocásticos embebidos, uno no es observable (estados ocultos), pero puede ser observado mediante el otro proceso (secuencia de observaciones).

La Figura A.1 muestra el esquema de estos dos modelos:

- a Las cadenas de Markov, donde $o1$, $o2$ y $o3$ son estados directamente observables
- b Los HMM, donde $q1$, $q2$ y $q3$ representan los estados ocultos, mientras que $o1$, $o2$ y $o3$ son los estados observables.

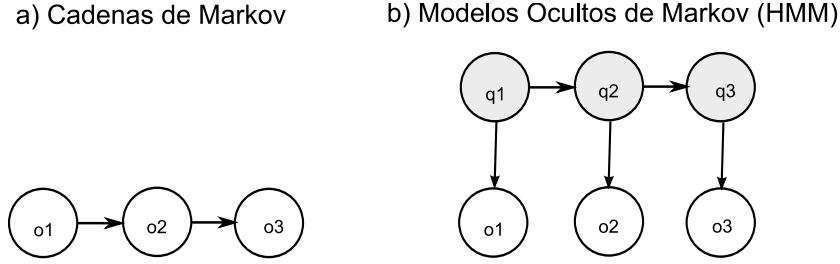


Figura A.1: Esquema de cadenas de Markov y Modelos Ocultos de Markov (HMM).

A.2. Formalización de los HMM

Para definir un HMM completamente, son necesarios los siguientes elementos:

- Espacio de estados del sistema: Consta de un conjunto de N estados distintos $S = \{S_1, S_2 \dots S_N\}$. Cada estado corresponde con un evento no observable.
- Un conjunto de parámetros $\theta = \{\pi, \mathbf{A}, \mathbf{B}\}$.
 - Probabilidad incondicional de estado: Es la probabilidad de un estado a priori, es decir, define la probabilidad de que S_i sea el primer estado de una secuencia. Los valores de sus parámetros se almacenan en el vector de estado inicial π y se representan $p(q_1 = i) \equiv \pi_i$.
 - Probabilidad condicional de transición: Define la probabilidad de ir del estado i al estado j . Se almacenan en la matriz \mathbf{A} y se representa $p(q_t | q_{t-1} \equiv a_{ij})$.
 - Probabilidad condicional de las observaciones o de emisión: Esta probabilidad caracteriza la verosimilitud de que el estado q_t genere una observación o_t . En el caso de esta tesis doctoral, donde las observaciones están definidas por un conjunto de valores discretos $O = \{o_k\}$, $k = 1, \dots, M$, la probabilidad de observar o_k dado el estado q_i se representa $b_i(k) = p(o_k | q_i) \equiv b_{ik}$.

Resumiendo, un HMM es un modelo probabilístico temporal en donde cada estado del proceso se describe por una o más variables estocásticas. Por ser un conjunto de variables aleatorias, sus posibles valores dependen una probabilidad de distribución. En cada instante discreto de tiempo el proceso asume estar en un estado, y al recibir las observaciones el proceso cambia de estado según su matriz de probabilidades de transición.

A.3. Los problemas fundamentales de un HMM

Rabiner [197] introdujo la idea de que los HMM deben caracterizarse por tres problemas fundamentales:

- **Problema1: Evaluación.** Dada una secuencia de observaciones $O = \{o_1, o_2 \dots o_T\}$ y un modelo $\theta = \{\pi, A, B\}$, este problema representa cómo determinar la verosimilitud $p(O|\theta)$ de dicha secuencia de observaciones dado el modelo.
- **Problema2: Descifrado.** Dada una secuencia de observaciones $O = \{o_1, o_2 \dots o_T\}$ y un modelo $\theta = \{\pi, A, B\}$, este problema representa cómo determinar la secuencia de estados más probable $Q = \{q_1, q_2 \dots q_T\}$, es decir, los estados que mejor expliquen las observaciones.
- **Problema3: Aprendizaje.** Dada una secuencia de observaciones $O = \{o_1, o_2 \dots o_T\}$ y una secuencia de estados $Q = \{q_1, q_2 \dots q_T\}$, aprender los parámetros del modelo $\theta = \{\pi, A, B\}$ que maximicen $p(O|\theta)$.

A.4. Verosimilitud de una secuencia de observaciones

La verosimilitud de una secuencia de observaciones $O = \{o_1, o_2 \dots o_T\}$ con respecto a un modelo $\theta = \{\pi, A, B\}$ se define según la ecuación A.17.

$$p(O|\theta) = \sum_{Q_i} p(O, Q|\theta) \quad (\text{A.1})$$

Dicha verosimilitud es la suma de la probabilidad condicional de la secuencia calculada sobre todas las posibles secuencias de estados generadas por el modelo. Y la probabilidad condicional dada una secuencia de observaciones $O = \{o_1, o_2 \dots o_T\}$ y una secuencia de estados $Q = \{q_1, q_2 \dots q_T\}$ se calcula tal y como se indica en la ecuación A.2.

$$p(O, Q|\theta) = p(O|Q, \theta)p(Q|\theta) \quad (\text{A.2})$$

Este cálculo es posible debido a la simplificación que caracteriza a las dos propiedades de Markov, a saber:

- **Propiedad1: Condición de Markov.** La condición de Markov establece que para una secuencia $Q = \{q_1, q_2 \dots q_T\}$, la probabilidad de alcanzar un nuevo estado sólo depende del estado anterior.
- **Propiedad2: Independencia de las observaciones.** La segunda propiedad considera que las observaciones son independientes con respecto al tiempo y que la probabilidad de una observación o_t únicamente depende del estado q_t que la ha generado.

Aplicando ambas hipótesis se obtiene que la estimación de la secuencia más probable dado un modelo se define:

$$\hat{\theta}_{MAP} = \underset{\theta_i}{\operatorname{argmax}} \sum_{Q_i} p(O|\theta) = \underset{\theta_i}{\operatorname{argmax}} p(q_1) p(o_1|q_1) \prod_{t=2}^T p(q_t|q_{t-1}) p(o_t|q_t) \quad (\text{A.3})$$

Ésta es la regla para estimar la secuencia de observaciones mediante la máxima verosimilitud.

A.4.1. Cálculo de la verosimilitud: El algoritmo *forward*

El algoritmo recursivo *forward* [197] calcula la probabilidad de que una secuencia observada (O) pertenezca a un determinado HMM (θ). Para ello, este algoritmo define una variable $\alpha_t(i)$ que se corresponde a:

$$\alpha_t(i) = p(o_1, o_2 \dots o_T, q_t = S_i|\theta) \quad (\text{A.4})$$

donde $\alpha_t(i)$ representa la probabilidad de haber observado la secuencia $O = \{o_1, o_2 \dots o_T\}$ y estar en el estado i en el instante t , dado el modelo θ . De este modo, la probabilidad de la secuencia de observaciones parcial es la suma de $\alpha_t(i)$ sobre todos los estados N .

El algoritmo recursivo *forward*, calculado sobre un modelo de N estados, consta de los siguientes pasos:

1. Inicialización:

$$\alpha_1(i) = \pi_i \cdot b_i(o_1), \quad 1 \leq i \leq N \quad (\text{A.5})$$

donde π_i es la probabilidad a priori de estar en el estado S_i en el instante de tiempo $t = 1$, y $b_i(o_1)$ representa la probabilidad de generar el símbolo o_1 en ese estado.

2. Recursión:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) \cdot a_{ij} \right] \cdot b_j(o_{t+1}), \quad 1 \leq t \leq T, \quad 1 \leq j \leq N \quad (\text{A.6})$$

3. Finalización:

$$p(O|\theta) = \sum_{i=1}^N \alpha_T(i) \quad (\text{A.7})$$

En este último paso, es decir, cuando se ha alcanzado el final de la secuencia observada, se suman las probabilidades de todos los caminos que convergen en el estado final N .

A.5. La secuencia de estados óptima: El algoritmo de Viterbi

Para un modelo que contiene variables ocultas, como es el caso del HMM, la tarea de determinar la secuencia de variables que genera la secuencia de observaciones se denomina tarea de decodificación o inferencia.

Un buen criterio de optimalidad consiste en seleccionar la secuencia de estados o camino que proporciona la máxima verosimilitud con respecto a un modelo dado. Esta secuencia puede obtenerse de manera recursiva mediante el algoritmo de *Viterbi*.

El algoritmo de *Viterbi* es la solución más popular para llevar a cabo el proceso de inferencia. Este algoritmo se compone de dos variables:

1. La máxima verosimilitud $\delta_t(i)$ a lo largo de una única secuencia entre todas las posibles secuencias que terminan en el estado i , en el instante de tiempo t .

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} p(q_1, q_2, \dots, q_t = S_i, o_1, o_2 \dots o_T | \theta) \quad (\text{A.8})$$

2. Una variable $\Psi_t(i)$ que permite mantener una memoria del “mejor camino” que termina en el estado i , en el instante de tiempo t .

$$\Psi_t(i) = \underset{q_1, q_2, \dots, q_{t-1}}{\operatorname{argmax}} p(q_1, q_2, \dots, q_t = S_i, o_1, o_2 \dots o_T | \theta) \quad (\text{A.9})$$

Ambas variables son vectores de N elementos, siendo N el número de estados generadores de observaciones.

Empleando estas dos variables, el algoritmo de *Viterbi* realiza los siguientes pasos:

1. Inicialización:

$$\delta_1(i) = \pi_i \cdot b_i(o_1), 1 \leq i \leq N, \Psi_1(i) = 0 \quad (\text{A.10})$$

donde, π_i es la probabilidad a priori de estar en el estado S_i en el instante de tiempo $t = 1$ y $b_i(o_1)$, representa la probabilidad de generar el símbolo o_1 en dicho estado.

2. Recursión:

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) \cdot a_{ij}] \cdot b_j(o_t) = \operatorname{argmax}_{1 \leq i \leq N} [\delta_{t-1}(i) \cdot a_{ij}] \quad (\text{A.11})$$

$$2 \leq t \leq T, 1 \leq j \leq N, \Psi_t(j)$$

El camino óptimo se compone a su vez de subcaminos óptimos. Para ello se estima el camino de máxima verosimilitud, considerando el camino de mejor verosimilitud alcanzado un paso antes y todas las transiciones a partir de él; el resultado se multiplica por la verosimilitud actual alcanzada en el estado actual. Por tanto el mejor camino se alcanza por inducción.

3. Finalización:

$$p^*(O|\theta) = \max_{1 \leq i \leq N} [\delta_T(i)] \cdot q_T^* = \operatorname{argmax}_{1 \leq i \leq N} [\delta_T(i)] \quad (\text{A.12})$$

La máxima verosimilitud se obtiene al alcanzar el final de la secuencia de observaciones.

4. Retropropagación:

$$Q^* = \{q_1^* \dots q_T^*\} q_t = \Psi_{t+1}(q_{t+1}^*), t = T-1, T-2, \dots, 1. \quad (\text{A.13})$$

La secuencia de estados se decodifica a partir de los vectores Ψ_t .

Por tanto, el algoritmo de *Viterbi* proporciona dos resultados de gran utilidad, dada una secuencia de observaciones $O = \{o_1, o_2 \dots o_T\}$ y un modelo $\theta = \{\pi, \mathbf{A}, \mathbf{B}\}$:

- La selección del camino $Q^* = \{q_1^*, q_2^* \dots q_T^*\}$ (de entre todos los posibles caminos generados por el modelo), que se corresponde con la secuencia de estados que proporciona la máxima verosimilitud para la secuencia de observaciones $O = \{o_1, o_2 \dots o_T\}$.
- La verosimilitud del mejor camino $p(O, Q^*|\theta) = p^*(O|\theta)$. Al contrario que el algoritmo *forward*, donde todos los posibles caminos son considerados, el algoritmo de *Viterbi* calcula la verosimilitud únicamente a lo largo del mejor camino.

A.6. El entrenamiento de los modelos: El algoritmo *Baum-Welch*

El tercer problema relacionado con los HMM es cómo aprender el conjunto de parámetros $\theta = \{\pi, A, B\}$ que definen el modelo. Formalmente, la solución consiste en estimar dichos parámetros, dada una secuencia de observaciones $O = \{o_1, o_2 \dots o_T\}$ y un conjunto de posibles estados.

El método clásico para el entrenamiento de los HMM es el algoritmo *Baum-Welch* o *forward-backward*, que es un caso particular del algoritmo de máxima expectación (Expectation-Maximization o algoritmo EM). Este algoritmo permite entrenar las probabilidades de transición **A** y las probabilidades de emisión **B** del modelo de forma conjunta.

A partir de una secuencia de observaciones, se calcula cuál es el camino que genera el modelo en función de los estados que han generado cada uno de los símbolos observados. Se obtiene una estimación de máxima verosimilitud de la probabilidad a_{ij} para cada transición concreta entre los estados i y j , contando el número de veces que tuvo lugar dicha transición ($C(i \rightarrow j)$), y después normalizando por el número total de veces que ocurre una transición desde el estado i :

$$a_{ij} = \frac{C(i \rightarrow j)}{\sum_{q \in Q} C(i \rightarrow q)} \quad (\text{A.14})$$

Sin embargo, no es posible realizar estos cálculos directamente a partir de las observaciones, ya que no es posible saber la secuencia de estados que genera el modelo, dada una entrada. Por ello, se realiza una estimación a través del cálculo de la probabilidad *forward* de una observación, para después dividir el total de la probabilidad entre los diferentes caminos que han contribuido a dicha probabilidad *forward*.

El siguiente paso en el método es aplicar el algoritmo *backward*, el cual representa la probabilidad de ver las observaciones desde el instante de tiempo $t + 1$

hasta el final de la secuencia, suponiendo que se está en el estado j en el instante t . La probabilidad *backward* β se define como se indica en la ecuación A.15.

$$\beta_t(i) = p(o_{t+1}, o_{t+2} \dots o_T \text{ vert } q_t = S_i, \theta) \quad (\text{A.15})$$

La probabilidad *backward* se estima mediante inducción de un modo similar al algoritmo *forward*.

1. Inicialización:

$$\beta_T(i) = 1, 1 \leq i \leq N \quad (\text{A.16})$$

2. Recursión:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} \cdot b_j(o_{t+1}) \cdot \beta_{t+1}(j), t = T-1, T-2, \dots, 1, 1 \leq i \leq N \quad (\text{A.17})$$

3. Finalización:

$$p(O|\theta) = \alpha_t(N) = \beta_T(1) = \sum_{j=1}^N a_{1j} \cdot b_j(o_1) \cdot \beta_1(j) \quad (\text{A.18})$$

Una vez se han estimado las probabilidades *forward* y *backward* puede procederse al cálculo de las probabilidades de transición a_{ij} y las probabilidades de observación $b_i(o_t)$ a partir de una secuencia de observaciones, aunque el camino tomado por el modelo sea oculto.

Para reestimar a_{ij} se calcula \hat{a}_{ij} tal y como se indica en la ecuación A.19.

$$\hat{a}_{ij} = \frac{\text{Número de transiciones esperadas desde el estado } i \text{ al estado } j}{\text{Número de transiciones esperadas desde el estado } i} = \frac{\sum_{t=1}^{T-1} \frac{\alpha_t(i) \cdot a_{ij} \cdot b_j(o_{t+1}) \cdot \beta_{t+1}(j)}{\alpha_T(N)}}{\sum_{t=1}^{T-1} \sum_{j=1}^N \frac{\alpha_t(i) \cdot a_{ij} \cdot b_j(o_{t+1}) \cdot \beta_{t+1}(j)}{\alpha_T(N)}} \quad (\text{A.19})$$

De forma análoga, se reestima la probabilidad de observación, que es la probabilidad de obtener la observación o_k , estando en el estado j . El valor de $\hat{b}_j(o_k)$ se estima tal y como se indica en la ecuación A.20.

$$\hat{b}_j(o_k) = \frac{\text{Número de veces esperadas de estar en el estado } j \text{ y observar el símbolo } o_k}{\text{Número de veces esperadas de estar en el estado } j} = \frac{\sum_{t=1}^T \sum_{s.t. o_t=o_k} \sum_{j=1}^N \frac{\alpha_t(i) \cdot a_{ij} \cdot b_j(o_{t+1}) \cdot \beta_{t+1}(j)}{\alpha_T(N)}}{\sum_{t=1}^T \sum_{j=1}^N \frac{\alpha_t(i) \cdot a_{ij} \cdot b_j(o_{t+1}) \cdot \beta_{t+1}(j)}{\alpha_T(N)}} \quad (\text{A.20})$$

*A.6. EL ENTRENAMIENTO DE LOS MODELOS: EL ALGORITMO BAUM-WELCH*179

Aplicando las ecuaciones A.19 y A.20 es posible reestimar la matriz de probabilidades de transición \mathbf{A} y la matriz de probabilidades de observación \mathbf{B} a partir de una secuencia O de observaciones.

Apéndice B

Meta Características por Vivienda

| Sensores | Baño - otros | Baño - puerta | Cocina - calentar | Cocina - almacenamiento | Puerta entrada casa | Dormitorio - otros | Dormitorio - puerta | Inodoro |
|---------------------------------|--------------|---------------|-------------------|-------------------------|---------------------|--------------------|---------------------|---------|
| Puerta de entrada | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Microondas | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Frigorífico | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Congelador | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Armario de la cocina (copas) | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Armario de la cocina (platos) | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Armario de la cocina (sartenes) | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Armario de la cocina (otros) | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Lavaplatos | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Lavadora | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Puerta del baño | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Puerta del lavabo | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Inodoro del lavabo | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Puerta de dormitorio | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |

Tabla B.1: Matriz de meta características para los datos del conjunto ‘KasterenA’. Representa qué meta características (columnas) se corresponden con qué sensores (filas).

| Sensores | Baño - otros | Baño - puerta | Cocina - calentar | Cocina - almacenamiento | Puerta entrada casa | Dormitorio - otros | Inodoro |
|----------------------------------|--------------|---------------|-------------------|-------------------------|---------------------|--------------------|---------|
| Puerta de entrada | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Puerta del balcón | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ventana | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Microondas | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Tostadora | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Frigorífico | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Fregadero de la cocina | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Armario de la cocina (otros) | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Armario de la cocina (platos) | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Armario de la cocina (cubiertos) | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Horno de la cocina | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Movimiento en la cocina | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Puerta del baño | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Movimiento en el baño | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Inodoro del baño | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Puerta de dormitorio | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Esterilla de la cama | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Esterilla de la cama | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Movimiento en el dormitorio | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Vestidor | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Esterilla del sofá | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Esterilla del sofá | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Tabla B.2: Matriz de meta características para los datos del conjunto ‘KasterenB’. Representa qué meta características (columnas) se corresponden con qué sensores (filas).

| Sensores | Baño - otros | Baño - puerta | Cocina - calentar | Cocina - almacenamiento | Puerta entrada casa | Dormitorio - otros | Inodoro |
|---------------------------------|--------------|---------------|-------------------|-------------------------|---------------------|--------------------|---------|
| Puerta de entrada | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Microondas | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Frigorífico | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Congelador | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Cajón de la cocina (otros) | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Cajón de la cocina (cubiertos) | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Armario de la cocina (sartenes) | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Armario de la cocina (otros) | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Armario de la cocina (platos) | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Armario de la cocina (copas) | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Puerta del baño1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Lavabo | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Movimiento en el baño | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Inodoro del baño1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Inodoro del baño2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Puerta del baño2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Puerta de dormitorio | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Vestidor | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Esterilla de la cama | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Esterilla de la cama | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Esterilla del sofá | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Tabla B.3: Matriz de meta características para los datos del conjunto ‘KasterenC’. Representa qué meta características (columnas) se corresponden con qué sensores (filas).

| Sensores | Ducha | Lavabo | Inodoro | Cama | Sofá | Cocina - almacenamiento | Cocina - calentar | Puerta entrada casa |
|----------------------|-------|--------|---------|------|------|-------------------------|-------------------|---------------------|
| Puerta de entrada | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Microondas | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Tostadora | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Vitrocerámica | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Frigorífico | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Alacena de la cocina | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Lavabo | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Inodoro | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Armario del baño | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ducha | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Esterilla de la cama | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Esterilla del sofá | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |

Tabla B.4: Matriz de meta características para los datos del conjunto ‘OrdonezA’. Representa qué meta características (columnas) se corresponden con qué sensores (filas).

| Sensores | Ducha | Lavabo | Inodoro | Cama | Sofá | Cocina - almacenamiento | Cocina - calentar | Puerta entrada casa |
|-----------------------|-------|--------|---------|------|------|-------------------------|-------------------|---------------------|
| Puerta de entrada | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Microondas | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Frigorífico | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Armario medicinas | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Puerta de la cocina | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Lavabo | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Inodoro | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Ducha | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Esterilla de la cama | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Puerta del dormitorio | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Esterilla del sofá | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Puerta del salón | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Tabla B.5: Matriz de meta características para los datos del conjunto ‘OrdenezB’. Representa qué meta características (columnas) se corresponden con qué sensores (filas).

| Sensores | Ducha | Lavabo | Inodoro | Cama | Sofá | Cocina - almacenamiento | Cocina - calentar | Puerta entrada casa |
|-----------------------|-------|--------|---------|------|------|-------------------------|-------------------|---------------------|
| Puerta de entrada | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Microondas | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Frigorífico | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Alacena de la cocina | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Puerta de la cocina | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Lavabo | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Inodoro | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Ducha | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Esterilla de la cama | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Puerta del dormitorio | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Esterilla del sofá | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Puerta del salón | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Tabla B.6: Matriz de meta características para los datos del conjunto ‘OrdonezC’. Representa qué meta características (columnas) se corresponden con qué sensores (filas).

Apéndice C

Evaluación de la Experiencia de Usuario

En el presente anexo mostramos los resultados de unas entrevistas que se han realizado con el objetivo de obtener una evaluación de nuestro sistema por parte de los voluntarios que han estado interaccionando con él.

El objetivo de dichas evaluaciones es estudiar el grado de aceptación que las tecnologías de monitorización empleadas consiguen por parte de los usuarios. De poco serviría nuestra propuesta si el entorno instrumentado que requerimos es rechazado por el público objetivo del sistema. Este estudio de la experiencia de usuario incluye una evaluación cuantitativa, que se ha realizado por medio de diferentes cuestionarios, y una evaluación cualitativa, que se ha llevado a cabo mediante entrevistas.

Las encuestas y entrevistas presentadas en este anexo nos ayudan a evaluar la satisfacción de los usuarios de la plataforma, haciendo especial hincapié en aspectos como la usabilidad, la intrusividad de los sensores y la sensación de privacidad. La población de este estudio se compone únicamente de tres usuarios y, por lo tanto, no define una muestra representativa. Sin embargo, aunque los resultados obtenidos no puedan ser considerados significativos, sí que son útiles para hacer una estimación preliminar acerca de cómo perciben el sistema sus futuros usuarios.

C.1. Evaluación Cuantitativa

Para elaborar los cuestionarios se ha empleado una escala de tipo Likert [238]. Esta escala es comúnmente utilizada en cuestionarios y encuestas de investigación, y define la métrica de evaluación mediante el nivel de acuerdo o desacuerdo de los usuarios con cada afirmación. Para nuestros tests hemos hecho uso de una

escala de siete puntos, donde el valor 1 representa el mayor nivel de conformidad y el valor 7 representa la mayor disconformidad. Estas pruebas están enfocadas a estimar diversos constructos o variables, concretamente la ansiedad, percepción de utilidad y percepción de intrusividad que el sistema genera, así como el esfuerzo que requiere por parte de los usuarios.

En la Tabla C.1 se muestran los resultados relativos al cuestionario que aborda la utilidad que perciben los usuarios. Tal y como se indicó en el Apartado 3.1.3, cada uno de nuestros voluntarios poseían un diferente nivel de autonomía, y esto queda así reflejado en la primera pregunta del formulario, que es la que obtiene una mayor disparidad de puntuaciones. Viendo el resto de respuestas se observa cómo los usuarios consideran este sistema muy útil de una forma mayoritaria. Nuestros voluntarios opinan que este tipo de tecnología les puede ser útil o muy útil para poder vivir de forma independiente durante más tiempo.

Tabla C.1: Evaluación de la percepción de utilidad del sistema.

| 01. Percepción de utilidad | Puntuación |
|--|------------|
| 1 Creo que soy capaz obtener ayuda autónoma y rápidamente si la necesito. | 3,0 |
| 2 Creo que si tengo algún accidente o caigo enfermo el sistema me será de ayuda. | 2,0 |
| 3 El sistema me permite vivir independientemente durante más tiempo. | 1,0 |
| 4 El sistema puede servir de ayuda a los cuidadores para ofrecer un mejor servicio. | 2,3 |
| 5 El sistema ha aumentado la percepción de seguridad en mi hogar, en comparación a cuando no lo tenía. | 2,3 |
| 6 De forma general, el sistema me parece útil. | 1,0 |

Los resultados relativos que evalúan la intrusividad del sistema se muestran en la Tabla C.2. Lo que más inquieta a nuestros voluntarios es la cuestión de cómo son tratados sus datos personales. Por otra parte, ninguna persona consideró en ningún momento que su intimidad estaba siendo vulnerada de algún modo, y consideraron que el sistema de sensores era completamente seguro para su salud.

El estudio de la ansiedad percibida por los usuarios se ha basado en la escala CARS (*Computer Anxiety Rating Scale*) [102]. Esta prueba está orientada a estimar el nivel de ansiedad que los sistemas tecnológicos pueden producir en las

Tabla C.2: Evaluación de la percepción de intrusividad del sistema.

| 02. Percepción de intrusividad/confianza | | Puntuación |
|--|---|------------|
| 1 | Considero que el sistema no vulnera mi intimidad. | 1,0 |
| 2 | Considero que mis datos son tratados y almacenados de forma segura. | 2,0 |
| 3 | Considero que mis datos no son empleados para otros propósitos ni cedidos a terceros. | 1,3 |
| 4 | Considero que el uso del sistema no supone ningún riesgo. | 1,3 |

personas. En nuestro caso, todos los voluntarios negaron que les produjese ansiedad alguna tener el sistema instalado en sus hogares. Observando los resultados de la Tabla C.3, se aprecia que el único punto que llegaba a preocupar a los voluntarios es el hecho de poder dañar algún componente de la red.

Tabla C.3: Evaluación de la ansiedad producida por del sistema.

| 03. Ansiedad producida por el sistema | | Puntuación |
|---------------------------------------|---|------------|
| 1 | Siento ansiedad al usar el sistema. | 7,0 |
| 2 | Me preocupa causar algún destrozo al sistema y provocar un mal funcionamiento. | 7,0 |
| 3 | Tengo problemas para entender los aspectos técnicos del sistema. | 4,6 |
| 4 | La instalación del sistema ha provocado cambios en mis hábitos de vida. | 7,0 |
| 5 | No sé si usar estas tecnologías por miedo ha cometer errores que no voy a poder corregir. | 5,0 |
| 6 | Evito utilizar tecnología porque me intimida y no me es familiar. | 6,6 |

Para elaborar el cuestionario relativo al esfuerzo que supone por parte del usuario tener este sistema desplegado en sus viviendas, hemos empleado el estudio Psicométrico del Índice de Carga Mental NASA-TLX [94]. Este procedi-

miento permite obtener una puntuación que defina la carga mental de una tarea. Su validez, comprobada en numerosas investigaciones, y su facilidad de uso hacen que sea el instrumento de evaluación de carga mental más ampliamente utilizado. El estudio se ha modificado para mantener la escala de siete puntos que hemos utilizado en el resto de la evaluación.

En lo referente al esfuerzo que supone por parte del usuario tener este sistema desplegado en sus viviendas, todos nuestros voluntarios realizaron las mismas valoraciones. Tal y como se aprecia en la Tabla C.4, todas las respuestas coinciden en que el nivel de esfuerzo que nuestro sistema requiere es el mínimo posible.

Tabla C.4: Esfuerzo requerido por el sistema.

| 04. Esfuerzo que necesita la utilización del sistema | | Puntuación |
|--|--|------------|
| 1 | Cuál es el nivel de esfuerzo mental requerido por el sistema. | 1,0 |
| 2 | Cuál es el nivel de esfuerzo físico requerido por el sistema. | 1,0 |
| 3 | Cuánto trabajo conlleva la interacción con el sistema. | 1,0 |
| 4 | Cómo de frustrante o estresante es interactuar con el sistema. | 1,0 |

La valoración general acerca del sistema se muestra en la Tabla C.5. Es importante esta valoración por parte de los usuarios, debido a que todos ellos mostraban deseos de volver a usar el sistema e incluso lo recomendarían encarecidamente a personas en su misma situación.

Tabla C.5: Intención de volver a usar el sistema.

| 05. Intención personal de volver a usar el sistema | | Puntuación |
|--|---|------------|
| 1 | Me gustaría volver a usar el sistema en el futuro. | 1,3 |
| 2 | Recomendaría el sistema a mis amigos o parientes que vivan solos. | 1,0 |

C.2. Evaluación Cualitativa

La evaluación cualitativa se llevó a cabo por medio de entrevistas personales. El objetivo de esta evaluación es obtener de primera mano la opinión de los voluntarios que han tomado parte en nuestros experimentos.

Estas evaluaciones constaron de diversas preguntas de carácter abierto acerca de los aspectos más positivos y los más problemáticos del sistema. Lo que los usuarios han valorado como algo especialmente positivo de la plataforma es su poca intrusividad. Todos los usuarios afirmaron que no fueron conscientes en ningún momento de estar siendo monitorizados, y el hecho de tener la red de sensores desplegada en sus viviendas no modificó en absoluto sus hábitos de vida.

Aquello que preocupaba más a los voluntarios fue que el sistema dejase de funcionar sin que pudiesen darse cuenta de ello. Bien porque a los sensores se les acabasen las baterías o por alguna acción directa suya, el hecho de no ser conscientes de cuándo el sistema estaba funcionando correctamente les hacía sentirse vulnerables.

Un aspecto a destacar es que nuestro voluntario con un mayor grado de dependencia fue el que mostró mayor interés en poder usar el sistema de forma real, después del experimento. La tecnología que dicha persona emplea actualmente para su supervisión es un botón de pánico, y la necesidad constante de llevar dicho dispositivo encima hacía aumentar su nivel de ansiedad.

C.3. Discusión

En este anexo de la tesis doctoral se ha evaluado cómo han percibido los voluntarios el sistema de modelado del comportamiento presentado. Mediante cuestionarios y entrevistas personales se ha podido determinar cómo de intrusivo para las personas es la red de sensores inalámbricos presentada y el nivel de utilidad de la propuesta.

Los aspectos más positivos del sistema según sus usuarios son la poca intrusividad que perciben por parte de la red de monitorización y el hecho de saberse controlados en todo momento, sin tener la necesidad de portar ningún otro dispositivo. Por otro lado, lo que más preocupó a nuestros usuarios es no saber si el sistema está funcionando correctamente, así como poder hacer algo que dañase o dejase inoperativa la red de sensores. El sistema puede considerarse exitoso desde el punto de vista de que los usuarios perciben que la plataforma aumenta su capacidad para mantener una vida independiente. Sin embargo, esta evaluación no puede considerarse significativa debido a la pequeña población de este estudio, por lo que una posible línea futura sería repetirla cuando el sistema se haya probado en un mayor número de viviendas.